

Open Research Online

The Open University's repository of research publications
and other research outputs

In vivo mapping of Polycomb and trithorax group proteins in chromatin of *Drosophila melanogaster*

Thesis

How to cite:

Strutt, Helen Lesley (1997). *In vivo* mapping of Polycomb and trithorax group proteins in chromatin of *Drosophila melanogaster*. PhD thesis. The Open University.

For guidance on citations see [FAQs](#).

© 1997 Helen Lesley Strutt

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Helen Lesley Strutt, M.A.

***In vivo* mapping of Polycomb
and trithorax group
proteins in chromatin of *Drosophila melanogaster***

A thesis in the biological sciences
submitted to the Open University
for the degree of Doctor of Philosophy

June 1997

② **Zentrum für Molekulare Biologie (ZMBH)
Universität Heidelberg**

and

① **National Institute for Medical Research
Mill Hill**

Date of award: 11th September 1997

ProQuest Number: C630550

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest C630550

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

- INDEX OF CONTENTS -

Contents	i
Tables and figures	v
Abbreviations	viii
Acknowledgements	xii
Abstract	xiii

- CONTENTS -

	page
Chapter 1	
Introduction	1
1.1 The selector gene hypothesis	1
1.2 The homeotic genes and the determination of segmental identity	2
1.3 Early pattern formation in <i>Drosophila</i>	4
1.4 Establishment of the expression patterns of homeotic genes	5
1.5 The Polycomb group genes	7
1.6 Molecular analysis of the Polycomb group genes	9
1.7 Open and repressive chromatin structures	12
1.8 Heterochromatin and position effect variegation in <i>Drosophila</i>	14
1.9 Analogies to yeast silencing paradigms	16
1.10 Polycomb group protein multimeric complexes	18
1.11 Polycomb group response elements	19
1.12 Models for Polycomb group protein function	21
1.13 Boundary elements in the bithorax complex	23
1.14 Genetic and molecular analysis of the trithorax group genes	25
1.15 Chromatin remodelling activities in <i>Drosophila</i>	26
1.16 The interaction between Polycomb and trithorax group gene products	29
1.17 Aims of the thesis	30
 Chapter 2	
Improvement of an <i>in vivo</i> formaldehyde cross-linking method for the analysis of protein-DNA interactions	32
2.1 Introduction	32

2.2	Comparison of the hybridisation of control and Polycomb-immunoprecipitated DNA to the bithorax complex walk	33
2.3	Effect of NdeII site distribution on PCR amplification efficiency	34
2.4	Blunt-end linker-modified PCR	35
2.5	Sonication efficiency of DNA from active and repressed genes	36
2.6	Determining the background level	38
2.7	Discussion	38
Chapter 3	Co-localisation of Polycomb protein and GAGA factor at regulatory elements of the bithorax complex	40
3.1	Introduction	40
3.2	Association of Polycomb protein with the bithorax complex	40
3.3	High resolution mapping of Polycomb immunoprecipitated DNA	42
3.4	Sequence comparison of Polycomb immunoprecipitated fragments	42
3.5	Binding of GAGA factor to consensus sequences in PREs	43
3.6	Binding of Polycomb and GAGA factor to the <i>Antennapedia</i> P1 promoter	44
3.7	Discussion	45
	<i>GAGA factor: a test of the cross-linking and immunoprecipitation method</i>	45
	<i>Association of Polycomb with PREs in the bithorax complex</i>	45
	<i>Polycomb protein spreads locally from PREs</i>	47
	<i>Binding of GAGA factor to DNA fragments containing PREs</i>	48
	<i>Role of GAGA factor at the Fab-7 boundary element</i>	50
Chapter 4	Association of Polycomb group proteins in a multimeric complex	51
4.1	Introduction	51
4.2	Purification of Polyhomeotic and Posterior sex combs fusion proteins	52
4.3	Production of Polyhomeotic and Posterior sex combs antibodies	53
4.4	Immunoprecipitation with Polyhomeotic and Posterior sex combs antibodies	55
4.5	Co-immunoprecipitation of Polycomb-lacZ fusion proteins	56
4.6	Discussion	57
	<i>The Polycomb group protein multimeric complex</i>	57

	<i>The role of the Polycomb chromodomain</i>	57
Chapter 5	The Polycomb group proteins form complexes of differential composition at different target genes	59
5.1	Introduction	59
5.2	Expression analysis of genes in the <i>engrailed-invected</i> region	59
5.3	Association of Polycomb protein with the <i>engrailed-invected</i> walk	60
5.4	Mapping of Polyhomeotic and Posterior sex combs binding sites in the <i>engrailed-invected</i> region	62
5.5	Regulation of gene VI expression by the Polycomb group proteins	64
5.6	Association of Polyhomeotic and Posterior sex combs with the <i>Abdominal-B</i> gene	64
5.7	Discussion	65
	<i>Polycomb binds to discrete elements in the regulatory regions of engrailed and invected</i>	65
	<i>Polycomb group proteins form multiple different complexes</i>	66
	<i>Association of Polyhomeotic and Posterior sex combs with active genes</i>	67
Chapter 6	Conclusions and outlook	69
6.1	The mechanism of silencing by Polycomb group proteins	69
6.2	The role of GAGA factor at distant regulatory elements	71
6.3	Multiple forms of Polycomb group protein complexes	71
Chapter 7	Materials and Methods	73
7.1	Materials	73
7.2	Immunoprecipitation of <i>in vivo</i> formaldehyde cross-linked chromatin from tissue culture cells	76
7.3	Purification of Polyhomeotic and Posterior sex combs fusion proteins	80
7.4	Purification of polyclonal antibodies against Polyhomeotic and Posterior sex combs proteins	82
7.5	Immunoprecipitation from nuclear extracts	83
7.6	Northern blot analysis of gene expression	85
7.7	<i>In situ</i> hybridisation to embryos	86

Chapter 8	References	89
Appendix		115
A1	Restriction endonuclease sites in the bithorax complex	115
A2	Restriction endonuclease sites in putative PREs of the bithorax complex	117
A3	Restriction endonuclease sites in the <i>engrailed-invected</i> genomic walk	120

- TABLES AND FIGURES -

		page
Table 1.1	The Polycomb group genes	11
Table 1.2	The cloned trithorax group genes	26
Table 3.1	Slot blot analysis of Polycomb and GAGA factor immunoprecipitated DNA in the bithorax complex	41
Table 4.1	Antibodies against Polyhomeotic and Posterior sex combs fusion proteins	54
Table 5.1	Slot blot analysis of Polycomb group and trithorax group protein immunoprecipitated DNA in <i>engrailed-inverted</i>	61
Table 5.2	Enrichment of Polycomb-associated PREs in the bithorax complex by Polyhomeotic and Posterior sex combs immunoprecipitations	63
Table 7.1	Subclones of the bithorax complex	75
Table 7.2	Subclones of the <i>engrailed-inverted</i> walk	75
Table 7.3	Subclones of the <i>Antennapedia</i> P1 promoter region	75
		page following
Figure 1.1	Genomic organisation and expression domains of the homeotic genes	2
Figure 1.2	Model for silencing at yeast telomeres	17
Figure 1.3	Association of Polycomb protein with inactive genes of the bithorax complex	22
Figure 1.4	Scheme of the formaldehyde cross-linking and chromatin immunoprecipitation method	30
Figure 2.1	Southern hybridisation of immunoprecipitated, PCR-amplified DNA to the bithorax complex genomic walk	33
Figure 2.2	Amplification of immunoprecipitated DNA is dependent on NdeII site distribution	34
Figure 2.3	Amplification of DNA using blunt-end linker-modified PCR	35
Figure 2.4	Southern hybridisation of control immunoprecipitated DNA to the bithorax complex genomic walk	36

Contents

Figure 2.5	Profile of hybridisation of control immunoprecipitated DNA to the bithorax complex	37
Figure 2.6	The efficiency of sonication of cross-linked DNA depends on gene activity	37
Figure 2.7	Analysis of plus and minus antibody immunoprecipitated DNA by slot blot	38
Figure 3.1	Southern hybridisation of Polycomb and GAGA factor immunoprecipitated DNA to the bithorax complex genomic walk	40
Figure 3.2	Distribution of Polycomb and GAGA factor on the bithorax complex	41
Figure 3.3	High resolution mapping of Polycomb and GAGA factor binding at PREs in the bithorax complex	42
Figure 3.4	Distribution of Polycomb and GAGA factor at PREs	42
Figure 3.5	Binding of GAGA factor to heat shock genes	43
Figure 3.6	Binding of GAGA factor to <i>Fab-7</i>	43
Figure 3.7	Distribution of Polycomb and GAGA factor at the <i>Antennapedia</i> P1 promoter PRE	44
Figure 4.1	Cloning of <i>polyhomeotic</i> and <i>Posterior sex combs</i> cDNAs into expression vectors	52
Figure 4.2	Expression of Polyhomeotic and Posterior sex combs fusion proteins	53
Figure 4.3	Western analysis using Polyhomeotic and Posterior sex combs antibodies	54
Figure 4.4	Co-immunoprecipitation of Polycomb, Polyhomeotic and Posterior sex combs proteins	55
Figure 4.5	Polycomb-lacZ fusion protein constructs	56
Figure 4.6	Co-immunoprecipitation of Polycomb-lacZ fusion proteins from nuclear extracts of transgenic flies	56
Figure 5.1	Northern analysis of genes in the <i>engrailed-invected</i> genomic region and <i>Abdominal-B</i>	59
Figure 5.2	Southern hybridisation of Polycomb group protein and GAGA factor immunoprecipitated DNA to the <i>engrailed-invected</i> genomic walk	60

Figure 5.3	Hybridisation of control immunoprecipitated DNA to the <i>engrailed-invested</i> walk	60
Figure 5.4	Distribution of Polycomb group proteins and GAGA factor on <i>engrailed-invested</i>	61
Figure 5.5	High resolution mapping of proteins at putative PREs in <i>engrailed-invested</i>	62
Figure 5.6	Expression of gene VI in wild type and mutant embryos	64
Figure 5.7	Association of Polycomb group proteins with <i>Abdominal-B</i> in Schneider cells	65
Figure A1	Correlation of the Polycomb and GAGA factor binding elements with the genetic elements of the bithorax complex	117
Figure A2	Detailed restriction map of the <i>engrailed-invested</i> genomic walk	120

- ABBREVIATIONS -

General

ANT-C	Antennapedia complex
ARS	autonomously replicating sequence
ATP	adenosine 5'-triphosphate
BCIP	5-bromo-4-chloro-3-indolyl phosphate
bp	base pair
BSA	bovine serum albumin
BX-C	bithorax complex
β-gal	β-galactosidase
cDNA	complementary DNA
cm	centimetre
CNS	central nervous system
dH ₂ O	distilled water
dATP	deoxyadenosine 5'-triphosphate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol-bis(β-aminoethylether) N,N,N',N'-tetraacetic acid
g	gram/gravity
GuHCl	guanidine hydrochloride
HCl	hydrochloric acid
Hepes	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulphonic) acid
HP1	heterochromatin-associated protein 1
HSE	heat shock element
HSF	heat shock factor
HMG-1	high mobility group protein 1
IPTG	isopropyl β-D-thiogalactopyranoside
ISWI	imitation switch protein
kb	kilobase
KCl	potassium chloride

kD	kiloDalton
<i>lac Z</i>	<i>β-galactosidase</i> gene
LB	Luria-Bertani bacterial medium
LiCl	lithium chloride
M	molar
MAR	matrix attachment region
mg	milligram
MgCl ₂	magnesium chloride
MgSO ₄	magnesium sulphate
ml	millilitre
mM	millimolar
MOPS	3-(N-morpholino)propanesulphonic acid
mRNA	messenger ribonucleic acid
μCi	microcurie
μg	microgram
μl	microlitre
μM	micromolar
NaCl	sodium chloride
NaH ₂ PO ₄	monosodium phosphate
NaAc	sodium acetate
NaOH	sodium hydroxide
NBT	nitroblue tetrazolium salt
ng	nanogram
NiSO ₄	nickel sulphate
nmol	nanomole
NP-40	Nonidet P-40
NURF	nucleosome remodelling factor
O.D.	optical density
ORC	origin recognition complex
ORF	open reading frame
PBS	phosphate-buffered saline
PcG	Polycomb group
PCR	polymerase chain reaction
PEV	position effect variegation

pg	picogram
PMSF	phenyl methyl sulphonyl fluoride
PRE	Polycomb group response element
RNA	ribonucleic acid
RNase A	ribonuclease A
rpm	revolutions per minute
RSC	"remodel the structure of chromatin" complex
SDS	sodium lauryl sulphate
SIN	switch independent gene (yeast)
SIR	silent information regulator gene (yeast)
SNF	sucrose non-fermenting gene (yeast)
SSC	sodium citrate-buffered saline
SV40	Simian virus-40
SWI	mating type switching gene (yeast)
TFIID	transcription factor II D
TRE	Trithorax response element
Tris	tris(hydroxymethyl)aminomethane
trxG	trithorax group
Tween-20	polyoxyethylene sorbitan-20
U	units
UV	ultraviolet
V	volt
v/v	volume to volume
°C	degrees Celsius

Amino acids

A	alanine	M	methionine
C	cysteine	N	asparagine
D	aspartic acid	P	proline
E	glutamic acid	Q	glutamine
F	phenylalanine	R	arginine
G	glycine	S	serine
H	histidine	T	threonine
I	isoleucine	V	valine
K	lysine	W	tryptophan
L	leucine	Y	tyrosine

Drosophila genes and regulatory sequences

In the text, gene names are in italics, and protein names are in capitals

<i>abd-A</i>	<i>abdominal-A</i>	<i>Kr</i>	<i>Krüppel</i>
<i>Abd-B</i>	<i>Abdominal-B</i>	<i>lab</i>	<i>labial</i>
<i>abx</i>	<i>anterobithorax</i>	<i>l(3)mbt</i>	<i>lethal (3) malignant brain</i>
<i>Antp</i>	<i>Antennapedia</i>		<i>tumour</i>
<i>ash 1 and 2</i>	<i>absent, small and homeotic discs</i>	<i>Mcp</i>	<i>Miscadastral pigmentation</i>
	<i>1 and 2</i>	<i>mod (mdg4)</i>	<i>modifier of mdg4</i>
<i>Asx</i>	<i>Additional sex combs</i>	<i>mx</i>	<i>multi sex combs</i>
<i>bcd</i>	<i>bicoid</i>	<i>nos</i>	<i>nanos</i>
<i>brm</i>	<i>brahma</i>	<i>pb</i>	<i>proboscipedia</i>
<i>bw</i>	<i>brown</i>	<i>pbx</i>	<i>postbithorax</i>
<i>bx</i>	<i>bithorax</i>	<i>Pc</i>	<i>Polycomb</i>
<i>bxd</i>	<i>bithoraxoid</i>	<i>Pcl</i>	<i>Polycomb-like</i>
<i>cad</i>	<i>caudal</i>	<i>ph</i>	<i>polyhomeotic</i>
<i>Dfd</i>	<i>Deformed</i>	<i>pho</i>	<i>pleiohomeotic</i>
<i>ems</i>	<i>empty spiracles</i>		<i>(l(4)102EFc or l(4)29)</i>
<i>en</i>	<i>engrailed</i>	<i>Psc</i>	<i>Posterior sex combs</i>
<i>esc</i>	<i>extra sex combs</i>	<i>Sce</i>	<i>Sex comb extra</i>
<i>eve</i>	<i>even-skipped</i>	<i>Scm</i>	<i>Sex comb on midleg</i>
<i>exd</i>	<i>extradenticle</i>	<i>Scr</i>	<i>Sex combs reduced</i>
<i>E(Pc)</i>	<i>Enhancer of Polycomb</i>	<i>scs</i>	<i>specialised chromatin structure</i>
<i>E(var)</i>	<i>Enhancer of variegation</i>	<i>snr1</i>	<i>snf5-related 1</i>
<i>E(z)</i>	<i>Enhancer of zeste</i>	<i>su(Hw)</i>	<i>suppressor of Hairy wing</i>
	<i>(polycombeotic)</i>	<i>Su(var)</i>	<i>Suppressor of variegation</i>
<i>Fab-7</i>	<i>Frontabdominal-7</i>	<i>Su(z)2</i>	<i>Suppressor of zeste 2</i>
<i>ftz</i>	<i>fushi tarazu</i>	<i>sxc</i>	<i>super sex combs</i>
<i>gt</i>	<i>giant</i>	<i>tll</i>	<i>tailless</i>
<i>hb</i>	<i>hunchback</i>	<i>Trl</i>	<i>Trithoraxlike</i>
<i>hh</i>	<i>hedgehog</i>	<i>trx</i>	<i>trithorax</i>
<i>hsp</i>	<i>heat shock protein</i>	<i>Ubx</i>	<i>Ultrabithorax</i>
<i>iab2 - iab9</i>	<i>infraabdominal 2 -</i>	<i>w</i>	<i>white</i>
	<i>infraabdominal 9</i>	<i>wg</i>	<i>wingless</i>
<i>inv</i>	<i>invected</i>	<i>y</i>	<i>yellow</i>
<i>kni</i>	<i>knirps</i>		

- ACKNOWLEDGEMENTS -

I would like to thank Sabine Messmer for transgenic flies expressing the 1 - 390 PC-lacZ fusion protein, and J.-M. Dura for the *ph* mutant fly stocks. P1 clones covering the BX-C were kindly provided by Ursi Weber and the Berkeley *Drosophila* genome project, λ clones covering the *en* walk by J. Kassis and T. Kornberg, and the *Scr*, *ph* and *Psc* cDNAs by W. Gehring, H. Brock and V. Pirrotta respectively. GAGA factor antibodies were a gift from Peter Becker, and control antibodies against PH and PSC proteins were from H. Brock and M. van Lohuizen.

Many thanks to Renato for general supervision and advice throughout my work. Valerio deserves thanks, not only for technical assistance, but also for providing the basis of my work and for support in those bad moments. Giacomo is thanked for many, many, many discussions, especially the useful ones. Thanks also to Achim, Axel, Bodo, Dani, Frank, Gabi, Gerhard, Hartmut, Martina, Steffen and Ute for all the entertainment. I especially appreciated the frequent exposure to traditional German music, and the lessons in speaking the little known German dialect "Deutschos".

Finally, I thank Bob Johnson, Rob Kay, and other members of their labs, for the valuable training I received prior to my PhD. Last but not least, special thanks to David for bothering to read the thesis, sharing his opinions about my projects, encouragement etc.

- ABSTRACT -

The Polycomb group and trithorax group genes of *Drosophila* are required for maintaining the repressed or active expression state of many developmental regulatory genes, such as the homeotic genes, throughout development. The Polycomb group genes have been suggested to act by regulating higher order chromatin structure.

This work describes an improvement in a formaldehyde cross-linking and chromatin immunoprecipitation technique for analysing *in vivo* protein-DNA interactions in tissue culture cells. Using this method, Polycomb protein was found to be strongly associated with previously-identified Polycomb group response elements (PREs) in repressed genes of the bithorax complex. Polycomb does not cover entire chromosomal domains, but spreads over a few kilobases of DNA surrounding PREs.

GAGA factor/Trithorax-like, a member of the trithorax group, is also bound at those PREs which contain GAGA consensus binding sites. This suggests that GAGA factor binds constitutively to PREs in the bithorax complex, which also function as trithorax group response elements.

Finally, Polycomb, Polyhomeotic and Posterior sex combs proteins are shown to participate in a common multimeric complex, and to be associated with identical regulatory elements of the selector gene *engrailed* in tissue culture cells. These three proteins are however differentially distributed on regulatory sequences of the *engrailed*-related gene *invected*. This suggests that there may be multiple different Polycomb group protein complexes which function at different target sites. Furthermore, Polyhomeotic and Posterior Sex Combs are associated with expressed genes, suggesting that the inclusion of Polycomb protein in the complex at PREs is required for stable silencing.

These results give insight into the mechanism by which the Polycomb group proteins mediate silencing of their target genes, and how this silencing is antagonised by the trithorax group genes. In addition, they reveal the structural and functional diversity of Polycomb group protein complexes.

- CHAPTER 1 -

Introduction

1.1. The selector gene hypothesis

A central question in biology concerns how a small number of undifferentiated cells can give rise to the diverse complex structures of the adult. In the fruitfly *Drosophila*, this process is achieved by first dividing the body of the embryo into a repeated array of identical segments. These are the basic "building blocks" of development, that subsequently diversify to produce different structures in each segment. The segments are units of cell lineage, and the metameric structure of the adult is a direct consequence of the basic organisation of the embryo. Two groups of founder cells from each embryonic segment (parasegment) are allocated to make either an anterior or posterior compartment of each adult segment. All descendents of a group of founder cells will exclusively form one adult compartment, and no other cells contribute to it (Morata and Lawrence, 1975). As embryonic parasegments are out of phase with the adult segments, the two groups of founder cells contributed by each parasegment form the posterior compartment of one adult segment and the anterior compartment of the next (see figure 1.1a).

It was proposed that compartmental identity is under the control of a small number of genes known as "selector" genes. This selector gene hypothesis suggests that the products of the selector genes act cell autonomously, and control developmental pathways by specifically activating other "realisator" genes. These realisator genes then define morphogenetic cell properties during differentiation. The combination of selector genes expressed in a group of founder cells determines the type of compartment it will construct, and mutation of a selector gene leads to an entire compartment developing the adult structures appropriate for another (Morata and Lawrence, 1975; Garcia-Bellido, 1975; Garcia-Bellido, 1977). Central to this hypothesis is the idea that activation or repression of selector genes occurs once and remains clonally irreversible. Moreover, the products of these selector genes are required throughout development to maintain the developmental pathway and to ensure a stable state of determination. Therefore the active or repressed state of expression of selector genes must be stably and heritably maintained throughout many cell divisions.

The homeotic genes and *engrailed* (*en*) are selector genes responsible for determining the identity of particular compartments. The mechanism of selector gene function is best understood in the case of *en*, which is expressed in the anterior region of each embryonic

parasegment (figure 1.1a) and defines the posterior compartment of each segment (Lawrence and Morata, 1976). *en* appears to have two discrete functions: firstly it has an autonomous function in establishing posterior identity, by controlling the expression of posterior-determining genes (Zecca et al., 1995; Tabata et al., 1995). It also has a non-autonomous effect on neighbouring cells, mediated through the secreted *hedgehog* (*hh*) gene product. Inductive interactions between anterior and posterior compartment cells lead to the formation of specialised cells at compartment boundaries. These boundary regions then play an important role in the growth and patterning of cells within each compartment (Blair, 1995; Hidalgo, 1996; Perrimon, 1995).

1.2. The homeotic genes and the determination of segmental identity

The homeotic genes are selector genes responsible for determining the identity of parasegments, and thus the type of adult structure formed within each segment (Lewis, 1978; Kaufman et al., 1990; Peifer et al., 1987). Homeotic gene mutations cause segments, or parts of segments, to be transformed into structures normally found in a different segment. For example, loss of function mutations in the *Antennapedia* (*Antp*) locus cause transformation of leg structures into antennae (Struhl, 1981b).

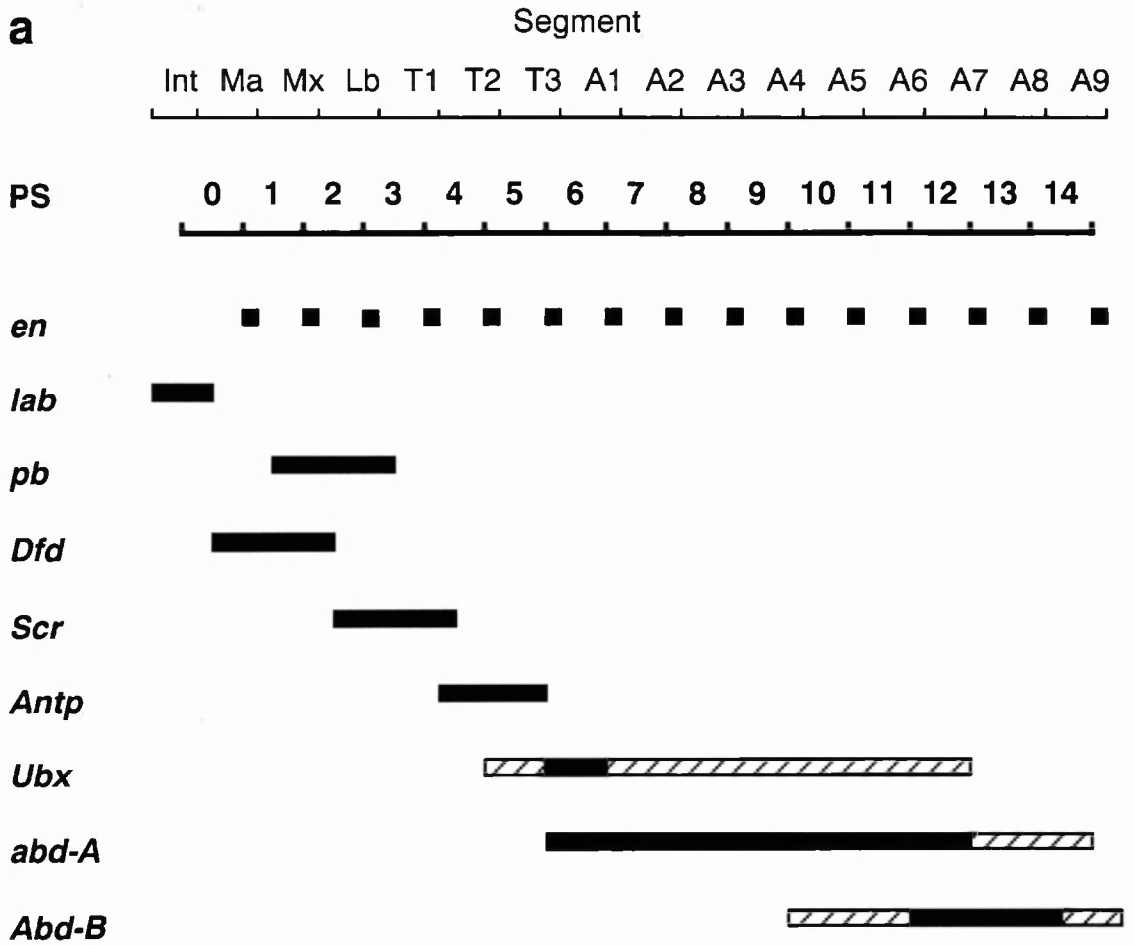
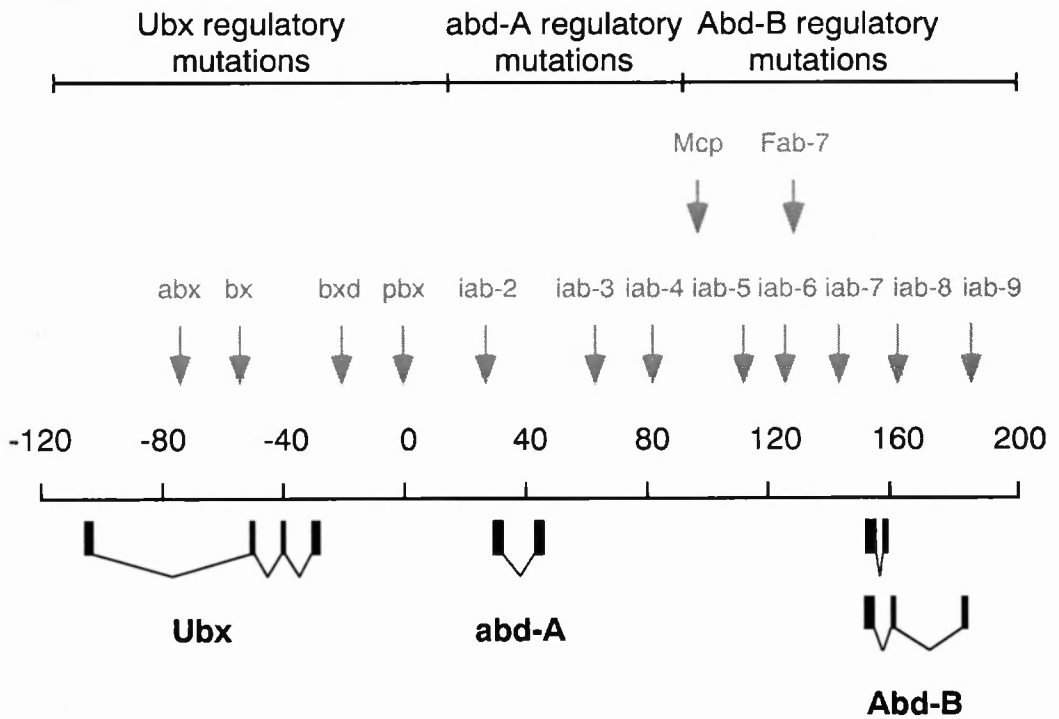
Homeotic genes are expressed in different subsets of parasegments in the embryo. They encode transcription factors that are presumed to regulate other genes which realise the segmental differences (McGinnis et al., 1984; Scott and Weiner, 1984). The gene or combination of genes expressed in a particular parasegment determines the segmental identity. The genes are located in two complexes on the third chromosome: the bithorax complex (BX-C) and the Antennapedia complex (ANT-C) (Sanchez-Herrero et al., 1985; Harding et al., 1985; Kaufman et al., 1990). They are arranged on the chromosome in the order in which they are expressed in the anterior-posterior axis of the animal. Most proximal on the chromosome is the *labial* (*lab*) gene of the ANT-C, which is expressed in the intercalary segment and required for the most anterior homeotic functions. Loss of function mutations in *lab* cause loss of head structures (Merrill et al., 1989). The *Abdominal-B* (*Abd-B*) gene of the bithorax complex is the most distal on the chromosome, and is expressed in the posterior abdominal segments of the embryo (see figure 1.1). The only exception to this rule of "colinearity" is the *proboscipedia* (*pb*) gene of the ANT-C (Pultz et al., 1988). However, this locus is unusual in that it is not required for embryonic patterning.

Interestingly, the organisation and expression of the homeotic gene complexes are conserved in vertebrates. For example there are four unlinked mouse homeobox (*Hox*) gene

Figure 1.1. Genomic organisation and expression domains of the homeotic genes.

(a) Expression domains of the homeotic genes in the epidermis. The segments of the adult fly are depicted at the top of the figure: Int (intercalary segment), Ma (mandibular segment), Mx (maxillary segment), Lb (labial segment), T1 - T3 (first thoracic to third thoracic segments), A1 - A9 (first abdominal to ninth abdominal segment). Below are the corresponding parasegments (PS) of the embryo. *en* is expressed in the anterior compartment of each parasegment, and thus defines the posterior compartment of each segment; cells not expressing *en* form the other compartment. Below are shown the appropriate expression domains in the epidermis of the homeotic genes of the ANT-C (*lab*, *pb*, *Dfd*, *Scr* and *Antp*) and the BX-C (*Ubx*, *abd-A* and *Abd-B*), listed in the order in which they are arranged on the chromosome. Black bars indicate regions of high expression, and striped bars regions of lower expression (Kaufman et al., 1990).

(b) Organisation of the BX-C. The coordinates of the BX-C (-120 to +200) are as previously described (Bender et al., 1983; Karch et al., 1985), and the positions of the *Ubx*, *abd-A* and *Abd-B* genes are shown at the bottom of the figure. Above are shown the positions of typical regulatory mutations affecting *Ubx* (*abx*, *bx*, *bx^d*, *pbx*), *abd-A* (*iab-2*, *iab-3*, *iab-4*) and *Abd-B* (*iab-5*, *iab-6*, *iab-7*, *iab-8*, *iab-9*). *Mcp* and *Fab-7* are unusual regulatory mutations which cause posterior transformations, and are required for regulation by the PcG gene products.

a**b**

clusters. Genes of these Hox complexes are arranged and expressed in the same relative order as their structural homologues in the *Drosophila* BX-C and ANT-C (McGinnis and Krumlauf, 1992; Krumlauf, 1994).

Loss of function mutations in the homeotic genes typically cause anterior transformations of segments (Lewis, 1978), suggesting the effect of a more anterior homeotic gene is only apparent in the absence of a posterior gene. Similarly, in mutant embryos in which all the homeotic genes are derepressed, all segments of the larval cuticle have a posterior abdomen phenotype (Lewis, 1978; Struhl, 1983). These results gave rise to the idea of phenotypic suppression, in which the activity of a more posterior gene product overrides that of a more anterior gene product (Lawrence and Morata, 1994). Transcriptional regulation may to some extent account for this posterior dominance, as more posterior gene products tend to repress transcription of the more anterior genes (see section 1.4). However there is also evidence that post-transcriptional events make a large contribution to this phenomenon (Lawrence and Morata, 1994).

The posterior thoracic segments and all nine of the abdominal parasegments are under the control of just three genes of the BX-C: *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*) and *Abd-B*. Each of these genes has specific regulatory regions, which regulate its expression in one parasegment. Therefore, *Ubx* has two regulatory domains (*abx/bx* and *bxd/pbx*) to control expression of *Ubx* in parasegments 5 and 6 respectively. Similarly, the regulatory domains *iab-2*, *iab-3* and *iab-4* control the expression of *abd-A* in parasegments 7, 8 and 9 (see figure 1.1b). Mutations in these domains cause anterior transformations of the single affected parasegment (Lewis, 1978). Interestingly, the principle of colinearity is preserved in these regulatory domains: the order of the domains on the chromosome corresponds to the order of the affected parasegments in the embryo. It was suggested that the order on the chromosome reflects a progressive opening of DNA domains; in parasegment 5 only the *abx/bx* enhancer is active, and progressively more regulatory domains become active in posterior regions of the embryo (Peifer et al., 1987). However, the functional relevance of colinearity is not clear, as the complex can still function if the *Ubx* locus is transposed to another chromosome (Struhl, 1984).

The domains of expression of the homeotic genes are determined by positional cues in the developing embryo (section 1.4). Thereafter, as the function of the homeotic genes is required continuously throughout development, an alternative mechanism is utilised to maintain their expression patterns. This mechanism involves the opposing functions of the products of the Polycomb group (PcG) and trithorax group (trxG) genes.

1.3. Early pattern formation in *Drosophila*

In order to understand the initial regulation of the homeotic genes it is necessary to describe briefly the processes of pattern formation in the *Drosophila* egg. Early patterning events are under the control of the maternal effect genes: genes which must be expressed in the mother for correct development of the embryo. These gene products both provide the embryo with polarity and lay down a pre-pattern for subsequent development (Ingham, 1988; St Johnston and Nüsslein-Volhard, 1992). The specification of the anterior-posterior axis is dependent on three determinant systems in the anterior, posterior and terminal domains of the body, whereas the Toll signalling pathway is sufficient for determining the dorsal-ventral axis. The anterior and posterior determinant systems depend on the localisation of maternally-encoded mRNAs in the oocyte (St Johnston and Nüsslein-Volhard, 1992), whereas the terminal and dorsal-ventral systems rely on the localised activation of uniformly-distributed cell surface receptors (Morisato and Anderson, 1995; Lu et al., 1993b). A common feature of all these processes however is that polarity is established before fertilisation, and results from signalling between the oocyte and the surrounding follicle cells of the mother (Ray and Schüpbach, 1996).

Based on the larval cuticular phenotypes of mutant embryos, it was proposed that three distinct classes of zygotic genes are required for the process of segmentation (Nüsslein-Volhard and Wieschaus, 1980). Firstly, the gap mutants lack contiguous blocks of segments. Secondly, the pair-rule mutants have only half the normal number of segments due to the deletion of alternate segments. Finally, in segment polarity mutants a specific section of each segment is affected. These genes act as a regulatory hierarchy that progressively subdivides the embryo into a repeating array of segmental units (Akam, 1987; Ingham, 1988).

The zygotic gap genes are expressed in broad, overlapping domains in the segmented part of the embryo. Their patterns of expression depend on the maternal effect genes responsible for anterior-posterior polarity, and in fact the gap genes are directly regulated by the transcription factors encoded by the maternal genes *bicoid* (*bcd*), *hunchback* (*hb*), *nanos* (*nos*) and *caudal* (*cad*) (Driever and Nüsslein-Volhard, 1988; Driever and Nüsslein-Volhard, 1989; Struhl et al., 1989; Gaul and Jäckle, 1987; Tautz, 1988; Rivera-Pomar et al., 1995). Furthermore, the spatial limits of gap gene expression domains are refined by interactions between adjacent gap gene domains (Jäckle et al., 1986; Pankratz et al., 1989).

Pair-rule genes are expressed in seven or eight stripes along the anterior-posterior axis of the embryo. They can be subdivided into two classes: the primary pair-rule genes, which derive their periodic pattern of expression from the maternal determinant systems and the gap genes; and the secondary pair-rule genes, which are dependent on the primary pair-rule genes for their expression (Ingham, 1988). Expression of primary pair-rule genes depends on a series of *cis*-acting elements in their promoters. Each element regulates expression of one stripe or of a subset of stripes, and contains a specific set of activator and repressor binding sites which are recognised by the maternal or gap gene transcription factors (Howard et al., 1988; Goto et al., 1989; Harding et al., 1989; Pankratz et al., 1990).

The expression of the segment polarity genes reveals the segmental organisation of the embryo. They are expressed in narrow stripes in each parasegment, and are responsible for defining different cell states within the parasegments. The selector gene *en* belongs to the segment polarity class of genes, and is present in a narrow stripe in the anterior of each parasegment (Fjose et al., 1985; Kornberg et al., 1985). This domain of expression is adjacent to that of another segment polarity gene *wingless* (*wg*) (Baker, 1987). Initial expression of the segment polarity genes is dependent on the transcription factors encoded by the pair-rule genes. When *en* expression is first apparent, the stripes alternate in intensity, suggesting a pair-rule modulation of early expression. Indeed, mutations in the pair-rule gene *fushi tarazu* (*ftz*) cause a loss of *en* expression in the even numbered parasegments (Martinez-Arias and White, 1988; Howard and Ingham, 1986).

Later in development, *en* expression is regulated by other segment polarity genes, in a process that involves cell-cell interactions between neighbouring *en* and *wg* expressing cells (DiNardo et al., 1988; Martinez-Arias et al., 1988). *en* expression then becomes independent of this extracellular influence, and relies on positive autoregulation (Heemskerk et al., 1991). However, this mode of *en* regulation is transient, and does not provide the mechanism for stable determination of the *en* cell fate (Heemskerk et al., 1991). This final mode of *en* regulation, like the homeotic genes, relies on the products of the PcG and *trxG* genes (Dura and Ingham, 1988; Busturia and Morata, 1988; Moazed and O'Farrell, 1992).

1.4. Establishment of the expression patterns of homeotic genes

The establishment of the initial domains of homeotic gene expression depends on the activity of transcription factors encoded by the gap and pair-rule genes. The *hb* gene product was shown to be important for establishing *Ubx* expression within its correct boundaries (White and Lehmann, 1986). *Ubx* is normally expressed between parasegments 5 and 13, but

is ectopically expressed both anteriorly and posteriorly to these boundaries in *hb* mutants. As the *hb* expression domains are approximately complementary to those of *Ubx*, it was proposed that *hb* directly represses *Ubx* expression. Similarly, the gap genes *Krüppel* (*Kr*), *knirps* (*kni*) and *tailless* (*tll*) have been implicated in regulating homeotic gene expression (Ingham et al., 1986; Harding and Levine, 1988; Irish et al., 1989; Casares and Sanchez-Herrero, 1995).

The first evidence that the pair-rule gene products regulate the homeotic genes came from the observation that the earliest expression of *Ubx* is in a transient pair-rule distribution, in frame with that of *ftz*, superimposed on a broader, continuous distribution of transcripts (Akam and Martinez-Arias, 1985). Indeed, some *ftz* alleles cause homeotic transformations of segments (Duncan, 1986), and *Ubx*, *Antp* and *Sex combs reduced* (*Scr*) were all found to be misexpressed in *ftz* mutants (Ingham and Martinez-Arias, 1986; Riley et al., 1987; Martinez-Arias and White, 1988). In general therefore, the pair-rule genes are required to activate selector genes in either even- or odd-numbered parasegments, whereas the gap genes determine the boundaries of homeotic gene expression. However, there are exceptions to this simple model, as *ftz* acts as a repressor of *Deformed* (*Dfd*) expression (Jack et al., 1988).

Once established, the patterns of expression of homeotic genes can be modulated by interactions with segment polarity genes. For example *Ubx* is expressed at lower levels in the posterior compartment of parasegment 6 than in the anterior compartment. This is due to repression by *en* in the anterior compartment (Martinez-Arias and White, 1988; Mann, 1994). In addition cross-regulation between homeotic genes plays a role in the final expression domains: for example, in *Ubx* mutants there is an increase in *Antp* expression in the posterior segments (Hafen et al., 1984). Posterior homeotic genes tend to repress more anterior genes; thus *Ubx* also represses *Scr* and *abd-A* and *Abd-B* repress *Ubx* expression (Struhl, 1982; Struhl and White, 1985). Finally, domains of expression may be stabilised by direct or indirect autoregulation, as has been shown for *Ubx*, *Dfd* and *lab* (Thüringer et al., 1993; Kuziora and McGinnis, 1988; Chouinard and Kaufman, 1991).

Reporter gene constructs were used to identify *cis*-acting regulatory sequences mediating expression of homeotic genes within their correct domains. The regulatory sequences of *Ubx* have been most extensively analysed, and a number of embryonic enhancers have been identified which direct expression of reporter genes in a pattern resembling that of *Ubx* (Simon et al., 1990; Irvine et al., 1991; Müller and Bienz, 1991; Qian et al., 1991; Christen and Bienz, 1994; Pirrotta et al., 1995). Each enhancer gave expression in either even- or odd-numbered pair-rule stripes, which respected a parasegment 5 or

parasegment 6 anterior boundary, depending on the parasegmental origin of the enhancer. Expression from some of these reporter constructs was lost in *ftz* mutants, consistent with *ftz* being a positive regulator of *Ubx* expression (Qian et al., 1991). Furthermore, these enhancer fragments contain binding sites for negative regulators such as the gap gene *hb* (Zhang et al., 1991; Qian et al., 1991; Pirrotta et al., 1995).

In many of these constructs the initial boundaries of expression were correct, but later in development ectopic expression was also seen in more anterior embryonic domains. This corresponds with the time that the transiently expressed products of the segmentation genes are lost from the embryo. Therefore these enhancer elements lack the information needed for maintaining the parasegmental pattern throughout later development. The ectopic expression of reporter constructs is similar to that observed for homeotic genes in embryos mutant for members of the PcG genes (section 1.5), and suggests that maintenance of the boundaries of selector gene expression later in development requires distinct regulatory elements.

1.5. The Polycomb group genes

Mutations in the *Polycomb* (*Pc*) gene cause posterior homeotic transformations. Flies heterozygous for *Pc* are viable, and show transformations of the second and third legs into first leg: this results in the appearance of extra sex combs on these legs (Lewis, 1978). Other dominant phenotypes resulting from *Pc* mutations include transformations of antennae into legs, of wings into halteres, and of ventral wing into dorsal wing (Lewis, 1978; Tjong and Russell, 1990). Homozygous *Pc* mutations cause embryonic lethality. In strong alleles, all thoracic and abdominal segments are transformed towards the eighth abdominal segment, and there is incomplete involution and dorsal close of the head (Lewis, 1978; Denell and Frederick, 1983). Homozygous mutations of a related gene *extra sex combs* (*esc*) do not cause embryonic lethality, but adult flies have extra sex combs on the second and third legs. Survival into adulthood is due to the persistence of the maternal product, which rescues the lack of zygotic *esc*. Indeed, homozygous *esc* mutant embryos from homozygous *esc* mutant mothers die at the end of embryogenesis, showing posterior transformations similar to those of *Pc* homozygotes (Struhl, 1981a). As these phenotypes are the opposite to those observed for loss of function mutations in the homeotic genes, it was suggested that *Pc* and *esc* are negative regulators of homeotic genes (Lewis, 1978; Struhl, 1981a).

Mutations in a number of other genes have been identified which cause similar phenotypes to those of *Pc* and *esc*. These genes are known collectively as the PcG genes (see

table 1.1; Jürgens, 1985). Mutations are embryonic lethal when homozygous, but in most cases embryos exhibit only partial posterior transformations of segments. However, embryos mutant for two or more PcG genes show homeotic transformations as strong as those seen in *Pc* homozygotes (Jürgens, 1985). Therefore, the PcG gene products may act synergistically to control the spatial regulation of BX-C genes. It was later shown that, as with *esc*, most PcG mutants are at least partially rescued by a maternal component. For example, *Sex comb extra* (*Sce*) or *Sex comb on midleg* (*Scm*) mutants from mothers carrying mutant germ-line clones show posteriorly-directed transformations as strong as those seen with *Pc* mutants (Breen and Duncan, 1986). However, the phenotypes of *Additional sex combs* (*Asx*), *Polycomb-like* (*Pcl*) and *super sex combs* (*sxc*) are still weaker than that of *Pc*, suggesting that some members of the PcG have a more important role in regulation of the homeotic genes than others (Breen and Duncan, 1986; Ingham, 1984).

As predicted, mutations in the PcG genes cause alterations in the expression patterns of homeotic genes. In *esc* mutants, the initial expression of *Ubx* is normal, and respects a parasegment 5 anterior boundary. Later in embryonic development however ectopic *Ubx* transcripts accumulate in all parasegments (Struhl and Akam, 1985). Finally, *Ubx* transcription decreases, presumably because of negative regulation by other indiscriminately-expressed homeotic genes. Similar ectopic expression of a variety of homeotic genes was observed in other PcG mutants, in embryonic and larval stages of development (Wedeen et al., 1986; Dura and Ingham, 1988; Glicksman and Brower, 1990; McKeon and Brock, 1991; Simon et al., 1992; Busturia and Morata, 1988). Because of phenotypic suppression effects, the ectopic expression of homeotic genes leads to posterior transformations (section 1.2).

The fact that homeotic gene expression is normal at early stages of development indicates that the PcG genes do not have a role in the establishment of the expression domains of these selector genes. However, they are required to maintain the repression of homeotic genes outside their normal boundaries at later stages of development. The timing of ectopic homeotic gene expression in PcG mutants suggests that their maintenance function is required when the initial repressor proteins encoded by the gap genes are no longer expressed.

The products of most PcG genes are required continuously throughout development. Interestingly, experiments using temperature sensitive alleles of *esc* have revealed that the *esc* gene product is only required during a small time window during embryogenesis (Struhl and Brower, 1982), a fact which explains the ability of maternal products to rescue

homozygous *esc* mutant embryos to adulthood. This would be consistent with *esc* having a role in the process of switching between gap gene-mediated repression and PcG-mediated repression.

With the exception of *esc*, the PcG genes are also required to maintain repression of the selector gene *en* (Dura and Ingham, 1988; Busturia and Morata, 1988; Moazed and O'Farrell, 1992). In addition, they appear to have a role in regulating segmentation genes of the gap and pair-rule classes. Firstly, mutations in *sxc*, *Pcl* and *pleiohomeotic* (*pho*) show segmentation defects resembling those of pair-rule genes or gap genes (Ingham, 1984; Breen and Duncan, 1986). Secondly, embryos which are doubly heterozygous for *polyhomeotic* (*ph*) and other PcG genes show gap, pair-rule and segment polarity segmentation defects. Thirdly, doubly heterozygous combinations of PcG and segmentation mutations enhance adult and embryonic segmentation defects (McKeon et al., 1994). Finally, both *Enhancer of zeste* (*E(z)*) and *pho* are required for maintaining the expression domains of the *kni* and *giant* (*gt*) gap genes (Pelegri and Lehmann, 1994).

Some PcG genes also appear to have more pleiotropic phenotypes. *ph* mutants were shown to have severe defects in the development of axon pathways in the embryonic central nervous system (CNS) (Smouse et al., 1988). Mutations in both *ph* and *Asx* cause ectopic expression of the *even-skipped* (*eve*) gene in cells of the CNS, together with a loss of *ftz* and *en* expression; this misexpression is likely to be the cause of the CNS defects seen in *ph* mutants (Dura and Ingham, 1988; Sinclair et al., 1992). Secondly, some PcG mutants may have regulatory functions in the germline: *multi sex combs* (*mxc*) mutants show defects in germ cell proliferation, and *E(z)* hypomorphic mutant females are sterile (Docquier et al., 1996; Phillips and Shearn, 1990).

These latter phenotypes do not appear to be associated with all PcG mutants. In addition, the strength of the segmentation defects caused by different PcG mutants is independent of the extent of their effects on homeotic gene expression. For example *sxc* shows segmentation defects which cannot be explained as a strong PcG phenotype, as *sxc* mutations do not cause the most severe homeotic phenotypes. Therefore, there may be a differing requirement for each PcG protein at different target genes, and some PcG proteins may have functions distinct from those of other PcG genes.

1.6. Molecular analysis of the Polycomb group genes

Thirteen genes causing PcG phenotypes have been identified (table 1.1), although it is estimated that 30 - 40 genes in total may be involved in PcG repression (Jürgens, 1985;

Landecker et al., 1994). The PcG proteins which have been molecularly characterised do not share significant sequence similarity. Some proteins however do contain conserved protein motifs which are seen in other nuclear proteins, and which are potential domains for protein-protein interactions. Firstly, the ESC protein contains a number of WD repeats (named because the motif usually ends with the amino acids WD) (Gutjahr et al., 1995; Simon et al., 1995; Sathe and Harte, 1995). These repeats are also found in the yeast repressor protein Tup1, where they mediate protein-protein contacts.

Secondly, both the SCM and PH proteins contain a SPM domain, a conserved sequence of 60 amino acids which is also seen in the product of the *lethal (3) malignant brain tumour (l(3)mbt)* gene (SPM is derived from: SCM, PH, MBT) (Bornemann et al., 1996). The structure of this domain is similar to that of helix-loop-helix domains, which mediate homodimeric or heterodimeric protein-protein interactions. Therefore it is likely that the SPM domain mediates homo- or heterodimerisation of SCM and PH. In addition the SCM protein contains two MBT repeats, 100 amino acid domains of unknown function which are shared between the *Scm* and *l(3)mbt* gene products (Bornemann et al., 1996).

A number of PcG proteins contain predicted zinc finger motifs. Both PH and SCM contain a single zinc finger motif, whereas Posterior sex combs (PSC) and PCL possess non-classical zinc finger motifs (the RING finger and PHD finger respectively) (DeCamillis et al., 1992; Bornemann et al., 1996; Brunk et al., 1991; van Lohuizen et al., 1991; Martin and Adler, 1993; Lonie et al., 1994). The RING finger motif in particular has been reported to show non-specific DNA binding, but no PcG protein (including PSC) has so far been demonstrated to bind DNA directly.

Molecular analysis of the PC protein gave an important insight into the mechanism of PcG repression. A 48 amino acid domain (the chromodomain) was identified which showed significant homology to the heterochromatin-associated protein HP1, encoded by the *Suppressor of variegation 205 (Su(var)205)* gene, and to the *Suppressor of variegation 3-9 (Su(var)3-9)* gene product (Paro and Hogness, 1991; James and Elgin, 1986; Tschiersch et al., 1994). As HP1 in particular is involved in heterochromatic silencing, it was proposed that the PcG proteins may silence target genes in a manner comparable to that of heterochromatin (Paro, 1990). A second link to heterochromatin emerged from the characterisation of the *E(z)* gene product: a 116 amino acid domain (the SET domain) was found in the products of both *E(z)* and *Su(var)3-9* (Jones and Gelbart, 1993; Tschiersch et al., 1994). Intriguingly, this domain is also found in the product of a *trxG* gene, *trithorax (trx)* (see also section 1.14).

Table 1.1. The Polycomb group genes

Gene	Cytological location	Conserved motifs	References
<i>Polycomb (Pc)</i>	78D	chromodomain	Paro and Hogness, 1991
<i>polyhomeotic (ph)</i>	2D	one zinc finger SPM domain	DeCamillis et al., 1992
<i>Posterior sex combs (Psc)</i>	49E	RING finger	Brunk et al., 1991 van Lohuizen et al., 1991 Martin and Adler, 1993
<i>Polycomb-like (Pcl)</i>	55A	PHD finger	Lonie et al., 1994
<i>Sex comb on midleg (Scm)</i>	85E	zinc fingers SPM domain MBT repeats	Bornemann et al., 1996
<i>Enhancer of zeste (E(z))</i>	67E	SET domain	Jones and Gelbart, 1993
<i>extra sex combs (esc)</i>	33B	WD repeats	Gutjahr et al., 1995 Sathe and Harte, 1995 Simon et al., 1995
<i>Additional sex combs (Asx)</i>	51A	not cloned	Jürgens, 1985 Breen and Duncan, 1986
<i>Sex comb extra (Sce)</i>	chromosome 3R, distal	not cloned	Breen and Duncan, 1986
<i>super sex combs (sxc)</i>	41C	not cloned	Ingham, 1984
<i>pleiohomeotic (pho)</i>	102EF	not cloned	Breen and Duncan, 1986 Girton and Jeon, 1994
<i>multi sex combs (mxc)</i>	8D	not cloned	Santamaria and Randsholt, 1995
<i>Enhancer of Polycomb (E(Pc))</i>	48A	not cloned	Sato et al., 1983 Sato et al., 1984

Molecular analysis has also led to the the identification of a number of mammalian homologues of the PcG genes. In particular the *M33*, *Rae-28/Mph1*, *eed* and *Enx-1* genes were found to be homologues of *Pc*, *ph*, *esc* and *E(z)* respectively (Pearce et al., 1992; Nomura et al., 1994; Alkema et al., 1997; Hobert et al., 1996; Schumacher et al., 1996). Two mouse

homologues of *Psc* have been characterised, the *bmi-1* and *mel-18* tumour suppressor genes, and the *mel-18* gene product was shown to bind DNA (Kanno et al., 1995). Mutations in *M33*, *eed*, *bmi-1* and *mel-18* cause posterior transformations of the axial skeleton (Coré et al., 1997; Schumacher et al., 1996; van der Lugt et al., 1994; Akasaka et al., 1996), whereas overexpression of *bmi-1* causes anterior transformations (Alkema et al., 1995). Furthermore, the *M33* gene is able to substitute for *Pc* in transgenic flies (Müller et al., 1995). Taken together, these results suggest a conservation in the function of the PcG genes, and thus in the genetic and molecular mechanisms which maintain the silenced state of homeotic genes.

1.7. Open and repressive chromatin structures

The shared homology between PC and HP1 proteins led to the proposition that PC, like HP1, is involved in regulating chromatin structure (Paro, 1990). In order to understand how PC/HP1 may function, this section will first discuss in more general terms the relationship of chromatin structure to gene expression.

The structure of chromatin was probed by the treatment of whole nuclei with micrococcal nuclease. Gel electrophoresis of the resulting DNA reveals a ladder of DNA fragments, with a periodicity of approximately 200 bp. More prolonged digestion gives a similar result, but this time the periodicity is 146 bp. This nuclease-resistant DNA is a component of the nucleosomal core particle: it is wrapped in approximately 2 superhelical turns over the surface of a histone octamer (2 copies of each of the highly basic histone proteins H2A, H2B, H3 and H4). The 50 bp of linker DNA between histones is more accessible to nuclease digestion, and interacts with the linker histone H1. This structure can also be visualised by electron microscopy of extended chromatin fibres at low ionic strength: a "beads on a string" structure is observed (van Holde et al., 1995).

Electron microscopy of chromatin fibres at higher salt concentrations revealed the presence of higher order chromatin structures. A thick fibre of 30 nm was observed, with beaded structures still retained within it. This fibre may result from the arrangement of nucleosomes in a regular helical array. Alternatively, it has been suggested that the 30 nm fibre is a less structured entity, and is the result of a linear array of clustered nucleosomes (van Holde et al., 1995).

Finally, electron micrographs of lampbrush chromosomes show that the chromatin fibre is organised into a series of loops emanating from the main axis of the chromosome. Current models suggests the subdivision of interphase chromosomes into a series of independent structural domains, which are anchored to the nuclear matrix at DNA elements

known as matrix or scaffold attachment regions (MARs or SARs) (Schedl and Grosveld, 1995).

Chromatin structure is an important factor in determining the transcriptional potential of a promoter, as the more chromatin is compacted, the less accessible it is to transcription factors. In fact, nucleosomes alone have a strong repressive effect on transcription, as many transcription factors are sterically hindered from interacting with their binding sites if the DNA is associated with a nucleosome. To bypass the repressive effects of nucleosomes, active promoters and enhancers usually possess a characteristic "open" chromatin configuration, which is more accessible to nuclease digestion than the bulk chromatin. Such hypersensitive sites result from the deposition of nucleosomes at precise positions on the surrounding DNA after DNA replication, and may be associated with non-histone chromosomal proteins (Becker, 1994; Wallrath et al., 1994). The promoters of other genes however, packaged in a continuous array of nucleosomes, require a perturbation or remodelling of chromatin structure for transcription factors to access DNA (see also section 1.15).

The involvement of histone acetyltransferases in remodelling higher order chromatin structure has recently been proposed (Tsukiyama and Wu, 1997). An open chromatin configuration is associated with increased levels of histone acetylation; conversely histones are generally under-acetylated at silent domains. Acetylation occurs at specific lysine residues in the N-terminal tails of histones, that protrude from the surface of the nucleosomes. This tends to weaken interactions between the histone tail and DNA, and whilst the nucleosomal core particle remains intact, the higher-order folding of nucleosomal arrays may be inhibited (Tsukiyama and Wu, 1997). A number of transcriptional activators have recently been found to be associated with histone acetyltransferases (Bannister and Kouzarides, 1996; Brownell et al., 1996; Mizzen et al., 1996; Ogryzko et al., 1996). However, the correlation between acetylation and gene activity is not perfect: deletion of the *Drosophila* deacetylase RPD3 leads to increased silencing, the opposite effect (De Robertis et al., 1996).

In *Drosophila*, the idea that transcriptional repression is associated with higher order chromatin structures arose from the cytological appearance of heterochromatin. The term heterochromatin is used to describe the chromosomal regions, primarily at centromeres, that remain condensed in interphase. Heterochromatic DNA is highly repetitive, largely transcriptionally inactive and late replicating (Eissenberg et al., 1995). Studies of *Drosophila* heterochromatin (section 1.8) and of the mechanistically-related silencing at yeast telomeres

(section 1.9), have led to insights into the nature of these higher order structures, and are of direct relevance in understanding silencing mediated by the PcG proteins.

1.8. Heterochromatin and position effect variegation in *Drosophila*

Position effect variegation (PEV) describes the inactivation of gene expression which occurs when a chromosomal rearrangement places a normally euchromatic gene in the vicinity of centromeric heterochromatin. Inactivation occurs in a mosaic fashion, and is accompanied by a cytologically visible spreading of heterochromatin into the rearranged euchromatic DNA. For example a rearranged *white (w)* gene is expressed in some of the cells in which it is normally active, but not in others, resulting in clonal patches of *w* expressing cells in the adult eye. This mosaic expression is thought to be the result of a stochastic inactivation process, depending on the degree of heterochromatin spreading; this inactivation event is then maintained throughout later development (Reuter and Spierer, 1992). This model assumes that there are elements in heterochromatin that promote spreading, and elements in euchromatin that terminate spreading; however, no such elements have yet been identified.

A large number of mutations have been identified which either enhance or suppress PEV. Suppressors of variegation (Su(var)) convert the mottled *w* eye phenotype to a more wild type colour. The wild type gene products are thus likely to promote the formation of heterochromatin, either because they are structural components of heterochromatin, or because they are enzymes that post-translationally modify heterochromatin proteins. Enhancers of variegation (E(var)) on the other hand cause a more mutant eye phenotype, and the wild type gene products are likely to inhibit the formation or spreading of heterochromatin. In addition, some modifiers of PEV show haplo-triplo effects: a reduction in the amount of gene product (i.e. 1 gene copy) may suppress variegation, whereas an increase in the same gene product (i.e. three gene copies) enhances variegation, or vice versa (Reuter and Spierer, 1992).

To explain the sensitivity of PEV to changes in the concentration of several modifier loci, a model was proposed in which the assembly of heterochromatin domains obeys the laws of mass action (Locke et al., 1988). In this model, gene inactivation by PEV is caused by the progressive and ordered assembly, along the chromosome, of the protein modules which form heterochromatin. Each module consists of all the protein subunits encoded by the modifier genes. By the law of mass action, a change in the solute concentration of each member of the complex will affect the concentration of the complex, and therefore the extent

of spreading along the chromosome and the degree of variegation (discussed in Henikoff, 1996). However, this may only hold true for those genes which show haplo-triplo-effects, such as the HP1 protein. Indeed, HP1 is localised on heterochromatic regions of polytene chromosomes, confirming that it may be a structural component of heterochromatin (James and Elgin, 1986; Eissenberg et al., 1990).

PEV was demonstrated to be associated with an altered chromatin structure. A transgene inserted in heterochromatin shows a more regular nucleosomal array, and is less accessible to restriction enzyme digestion in isolated nuclei than a transgene inserted at a euchromatic site (Wallrath and Elgin, 1995). These results suggest that DNA is more compacted in heterochromatin. However, little difference in accessibility to an *E. coli* dam methylase gene was observed between euchromatic and heterochromatic target sites (Wines et al., 1996).

An alternative model for the mechanism of heterochromatin silencing has recently been proposed, which does not rely on a linear propagation of heterochromatin subunits along the chromosome. Repetitive arrays of the *w* gene in euchromatic sites were shown to exhibit PEV phenotypes (Dorer and Henikoff, 1994). Variegation is stronger when the repeat sequences are in the vicinity of heterochromatin, and is sensitive to modifiers of PEV. The degree of variegation increases with increasing copy number of repeats, and it was suggested that pairing of repetitive sequences creates a structure that can be recognised by heterochromatin proteins. Furthermore, PEV of a *brown* (*bw*) transgene is enhanced when the transgene is locally duplicated, the degree of enhancement depending on both copy number and orientation (Sabl and Henikoff, 1996). These results argue against a linear propagation model for heterochromatin, which should not be influenced by DNA orientation. Instead, a pairing-looping model for PEV was invoked, in which heterochromatin forms as a result of local pairing between homologous DNA sequences. Chromosomal rearrangements which move a gene closer to heterochromatin may lead to pairing of elements near the gene and similar sequences in heterochromatin (for example the middle repetitive elements clustered in heterochromatin and scattered throughout euchromatin). Inactivation is thus the result of the gene being brought closer to a heterochromatin compartment.

The idea that genes are silenced by specific contact with centromeric heterochromatin is supported by experiments using the *bw^D* allele, a dominant null mutation caused by the insertion of a large block of heterochromatin into the *bw* coding sequence. Fluorescence *in situ* hybridisation experiments showed that the heterochromatin insertion in *bw^D* physically associates with other heterochromatic regions of the same chromosome, in a stochastic

manner. Interestingly, a wild type *bw* allele (which normally pairs with the homologous *bw^D* allele) is also caused to associate with centromeric heterochromatin, and is inactivated (Csink and Henikoff, 1996; Dernburg et al., 1996).

In summary, two models have been proposed for the mechanistic basis of heterochromatin silencing. In both models proteins encoded by suppressors of PEV would form silencing complexes. In the first model these silencing complexes propagate linearly along the chromosome, and package DNA into a compact, inaccessible structure. An alternative model however suggests that heterochromatin is synthesised by the pairing of repeat arrays. DNA elements close to these paired repeat sequences may then be inactivated, as the presence of nearby heterochromatin is incompatible with transcription (Henikoff, 1994; Henikoff, 1996).

1.9. Analogies to yeast silencing paradigms

In the yeast *Saccharomyces cerevisiae* the expression of genes placed near to the ends of chromosomes is subject to position effects. The telomeres create heritably repressed states that extend continuously for varying distances (3 - 5 kb) along the chromosomes from the telomere. The extent of spreading varies stochastically in different cells and results in a variegated phenotype (Gottschling et al., 1990; Renauld et al., 1993). The similarity between telomeric position effects and PEV suggests that they utilise similar silencing mechanisms, and the use of yeast as a model system has provided many insights into how heterochromatin-like structures may be established.

Modifiers of position effect at the telomeres are shared between another silencing phenomenon in yeast, that of the silent mating type loci HML and HMR (Aparicio et al., 1991). The mating type genes *a* and α are expressed when present at the MAT locus, but transcriptionally silent when at HML and HMR, although all the cis-acting sequences required for expression are present. The HML and HMR loci are flanked by two silencer elements E and I, which are responsible for transcriptional repression. They contain three discrete binding elements. Two of these elements are bound by the ABF1 and RAP1 transcription factors respectively; the RAP1 factor also associates with poly (C₁-3A) sequences of telomeres. The third element is an autonomously replicating sequence (ARS), found at all yeast replication origins. The origin recognition complex (ORC) binds to the ARS of the HML and HMR silencer, where it is required for silencing as well as for replication (Laurenson and Rine, 1992; Pillus and Grunstein, 1995; Rivier and Pillus, 1994).

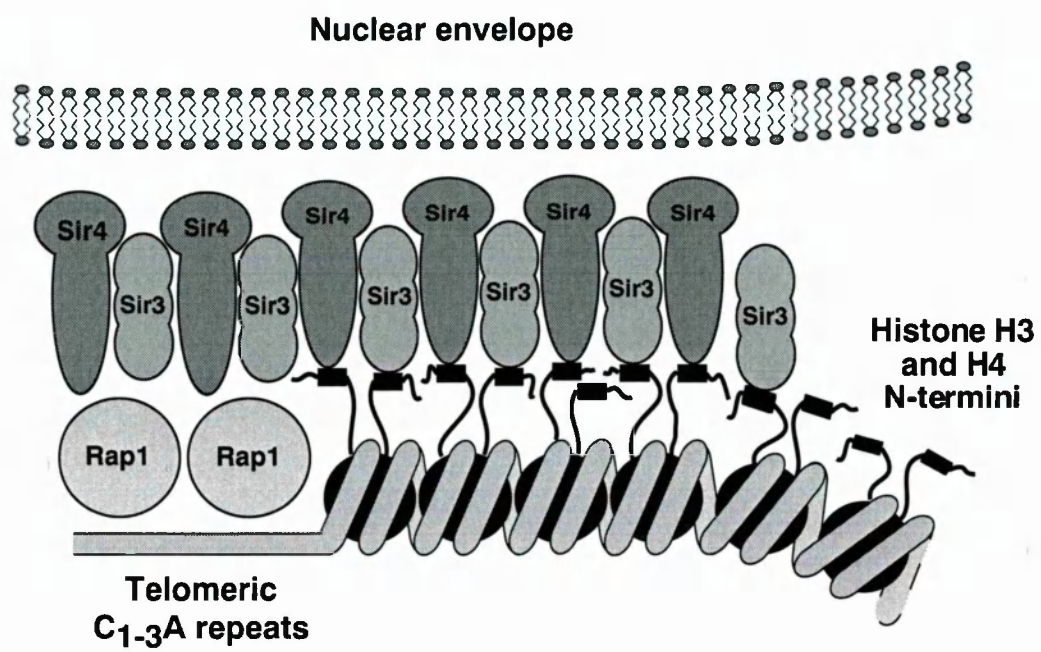
The non-DNA binding SIR (silent information regulator) proteins are essential for silencing both at telomeres and at the mating type loci (Aparicio et al., 1991). Mutations in SIR2, SIR3 and SIR4 abolish repression whereas mutations in SIR1 have less severe effects. A number of experiments have indicated that the SIR proteins play a structural role in the maintenance of silent chromatin. The involvement of SIR3 in such a structure was proposed by the finding that overexpression of SIR3 enhances telomeric position effects, allowing the repression of genes at increasing distances from the telomere (Renauld et al., 1993). SIR3 and SIR4 interact directly with the C-terminus of RAP1 (Moretti et al., 1994). In addition immunofluorescence studies show that RAP1, SIR3 and SIR4 are co-localised in foci formed by the association of telomeres at the nuclear periphery (Palladino et al., 1993). RAP1 was therefore proposed to recruit the SIR proteins to the chromosome. Silencing could also be established by artificially tethering the SIR1 protein to DNA; it was dependent on SIR2, SIR3 and SIR4 function, and therefore it was proposed that SIR1 is also important in recruitment of SIR protein complexes at the HM loci (Chien et al., 1993).

Interestingly, mutations in histones H3 and H4 derepress silencing both at the HM loci and at telomeres (Kayne et al., 1988; Thompson et al., 1994), suggesting a direct link between silencing and chromatin structure. In fact, SIR3 and SIR4 interact directly with the N-terminal tails of histones H3 and H4 (Hecht et al., 1995). Immunoprecipitation of either SIR2, SIR3 or SIR4 from formaldehyde cross-linked chromatin revealed that they are associated with DNA, spreading from the telomeres as far as silencing extends (Hecht et al., 1996; Strahl-Bolsinger et al., 1997). Therefore, a general model for telomeric silencing was proposed (figure 1.2), in which RAP1 recruits the SIR proteins to telomeres, and initiates the assembly of a multimeric complex. SIR3 and SIR4 form a heterochromatin-like structure that spreads along the chromosome, by direct interactions with histones. This structure would restrict the access of transcription factors to *cis*-regulatory elements. As SIR proteins have been observed to interact with each other (Chien et al., 1993; Hecht et al., 1996; Strahl-Bolsinger et al., 1997), it is likely that multiple histone-SIR and SIR-SIR interactions assist in the cooperative spreading of the heterochromatic state. Furthermore, mutation of any single SIR protein appears to destabilise the association of all of the other SIR proteins in silent chromatin (Strahl-Bolsinger et al., 1997).

Although the establishment of silencing occurs independently at telomeres and the HM loci, activity of the HM silencers is dependent on proximity to the telomeres (Maillet et al., 1996). SIR3 and SIR4 fusion proteins which are artificially tethered to DNA with a GAL4 DNA binding domain can bypass the requirement for a silencer and establish silencing of

Figure 1.2. Model for silencing at yeast telomeres.

RAP1 protein binds to C₁₋₃A repeat sequences of telomeric DNA, and in turn recruits the SIR2, SIR3 and SIR4 proteins to telomeres (SIR3 and SIR4 are shown). SIR3 and SIR4 interact with each other and with the N-termini of histones H3 and H4, and spread along the chromatin fibre. SIR3 and SIR4 are also required for the association of telomeres with the nuclear envelope. From Hecht et al., 1995.



reporter genes. This effect however only occurs if the construct is inserted near the telomeres (Marcand et al., 1996a). Telomeres may therefore boost silencing by creating a high local concentration of silencing factors in a particular nuclear compartment. The proximity of a gene to such a compartment may affect its susceptibility to silencing, as has also been observed for the *bw^D* allele of *Drosophila* (Marcand et al., 1996b).

1.10. Polycomb group protein multimeric complexes

The homology between the chromodomains of the PC and HP1 proteins led to the idea that PcG proteins may maintain the repressed state of target genes in a manner comparable to that of heterochromatin (Paro and Hogness, 1991; Paro, 1990). Double and triple mutant combinations of the PcG genes show a synergistic enhancement of the homeotic phenotypes observed with single mutants (Jürgens, 1985). These synergistic enhancements are consistent with the PcG proteins acting in a common regulatory structure. Furthermore, *ph* exhibits extragenic non-complementation with a number of other PcG genes, suggesting that these PcG gene products may interact directly with PH (Cheng et al., 1994).

A molecular interaction between a number of PcG proteins was confirmed by recent experiments. Firstly, both PC and PH were shown to be present in a multimeric protein complex which contains at least 10 - 15 other proteins (Franke et al., 1992). Mouse homologues of PC, PH and PSC are also constituents of a multimeric complex (Alkema et al., 1997), and PH from both mouse and human can homodimerise with itself and interact with PSC (Alkema et al., 1997; Gunster et al., 1997). In addition, the *Xenopus* homologues of PC and PSC were shown to be capable of interacting *in vitro* (Reijnen et al., 1995). Secondly, PC was shown to bind to approximately 100 specific sites on polytene chromosomes of larval salivary glands. Many of these binding sites correspond to the positions of PcG target genes, including the homeotic genes of the BX-C, ANT-C and *en* (Zink and Paro, 1989; Paro and Zink, 1992). Chromosomal binding appears to be mediated by the chromodomain, probably via association with other proteins (Messmer et al., 1992; Platero et al., 1995). Both PH and PCL bind to polytene chromosomes with a distribution identical to that of PC (Franke et al., 1992; Lonie et al., 1994), whereas PSC has overlapping but non-identical binding sites (Martin and Adler, 1993; Rastelli et al., 1993). Temperature sensitive mutations in *E(z)* cause a loss of PSC and PC binding to their polytene chromosome sites at the restrictive temperature (Rastelli et al., 1993; Platero et al., 1996). This result suggests that *E(Z)* and PC also bind to identical polytene chromosome sites, although *E(Z)* was only detected at a

subset of the PC sites (Carrington and Jones, 1996). Finally, a chimaeric protein, in which the chromodomain of HP1 is replaced by that of PC, is targeted both to the normal heterochromatic HP1 binding sites and to PC binding sites on polytene chromosomes. PSC is recruited to binding sites of the chimaeric protein in heterochromatin, suggesting that it interacts (not necessarily directly) with the PC chromodomain (Platero et al., 1995).

Although the PcG proteins appear to act together in a complex similar to heterochromatin components, there is little evidence for the members of this complex obeying the laws of mass action. The possible exceptions are *ph*, *Pcl* and *esc* which show haplo-enhancer, triplo-suppressor effects. For example, a duplication of the *Pcl* locus suppresses the extra sex comb phenotype of *Pc* heterozygotes (Kennison and Russell, 1987). In addition, *ph* duplications suppress the phenotype of *Pc* and *Pcl* mutants, and *esc* duplications suppress the *Pcl* extra sex combs phenotype (Cheng et al., 1994; Campbell et al., 1995). Interestingly, duplications of *esc* in an otherwise wild type background cause anterior transformations, the opposite phenotype to that of a loss-of function mutation (Campbell et al., 1995).

Increasing the dose of other Polycomb group genes either has no effect or actually enhances posterior transformations, an outcome opposite to that predicted by the law of mass action (Campbell et al., 1995). These results therefore argue against a simple mass action model for PcG protein function, in which one PcG protein can replace the function of other members of the complex. Instead, it is likely that PcG proteins have different, unique roles either in the formation or the function of the PcG silencing complex.

1.11. Polycomb group response elements

Cis-regulatory elements which can maintain transcriptionally-repressed states throughout development were identified in the BX-C (Müller and Bienz, 1991; Busturia and Bienz, 1993; Qian et al., 1993; Simon et al., 1993; Chan et al., 1994; Christen and Bienz, 1994), the ANT-C (Zink et al., 1991; Gindhart and Kaufmann, 1995) and the *en* locus (Kassis, 1994). These regulatory elements were termed PcG response elements (PREs). The use of reporter constructs in transgenic flies revealed two characteristic properties of PREs. Firstly, they are able to maintain the boundaries of the homeotic gene expression domains throughout development, in a manner dependent on the function of the PcG genes. Secondly, they are able to recruit PC protein to an ectopic site in polytene chromosomes, which corresponds to the position at which the transposon is inserted.

Interestingly, it was shown that PREs can induce variegated expression of reporter genes in the transposon construct (Fauvarque and Dura, 1993; Chan et al., 1994; Gindhart and Kaufmann, 1995; Zink and Paro, 1995). For example, the *w* gene, used as a selectable marker for the transformation of the reporter construct in flies, shows variegation of expression in the adult eye reminiscent of that seen in PEV. Thus, repression is clonal in nature and the distribution of *w*-expressing and non-expressing cells indicates that repression is established early in embryogenesis. Mutations in *Pc* and *ph* were shown to modify variegation. Therefore it was proposed that the binding of a complex of PcG proteins to PREs generates a localised heterochromatinisation which, by analogy to PEV, leads to the variegated phenotype (Fauvarque and Dura, 1993). It was also shown that mutations in some *Su(var)* genes affect variegation at PREs (Fauvarque and Dura, 1993; Chan et al., 1994), which supports the idea that the PcG proteins and the *Su(var)* proteins form related complexes that function by a similar mechanism.

The BX-C has been most extensively analysed for the presence of PREs. The entire *Ubx* region and most of *abd-A* was scanned by linking individual fragments to reporter constructs. One element with PRE activity was found in every parasegmental regulatory domain, indicating that one PRE is sufficient to control all the regulatory elements of that parasegment (Chiang et al., 1995). A model was proposed in which PcG proteins nucleate at PREs and spread in a heterochromatin-like fashion to stably repress entire parasegmental regulatory domains (Pirrotta and Rastelli, 1994; Paro, 1995). Such a spreading would be analogous to that proposed for the SIR protein complexes at the silent mating type loci and telomeres of yeast.

The nucleation of PcG proteins on DNA could be mimicked by expressing PC as a fusion protein with the DNA binding domain of the yeast GAL4 transactivator, which leads to the artificial tethering of PC at GAL4 binding sites (Müller, 1995). Neighbouring reporter genes are silenced, in a manner dependent on the endogenous PcG proteins, suggesting that a silencing complex is recruited by interaction with GAL4-PC. Interestingly, if GAL4-PC is provided transiently, silencing of reporter genes is maintained even when the expression of GAL4 is no longer induced. Therefore, the presence of endogenous PC proteins at the target site appears to be sufficient to maintain a permanent alteration in the chromatin structure. However, the permanent silencing of reporter genes was dependent on the type of sequence in the reporter construct (Müller, 1995). It was therefore proposed that spreading of a heterochromatin-like structure from a nucleation site at PREs was dependent on specific

maintenance elements. This spreading is necessary for the long-term stability of the silenced state (Paro, 1995).

Two models have been put forward to explain how the PcG proteins may be first nucleated at PREs of inactive domains (Bienz and Müller, 1995). Firstly, the finding that many *Ubx* embryonic enhancers contain binding sites for the gap gene repressor HB led to the hypothesis that such repressors interact with and recruit PcG proteins to PREs, and direct the assembly of a silencing complex. In support of this model, the presence of multiple HB binding sites in a reporter construct was sufficient to cause PcG-dependent silencing (Zhang and Bienz, 1992). However, constructs containing PREs but no HB binding sites can also establish repression and silence reporter genes (Chan et al., 1994). Therefore an alternative hypothesis suggests that the formation of a complex may depend on the state of activity of promoters or enhancers in the vicinity. For example formation of a silencing complex may be inhibited in genes which are active and in which enhancer-promoter looping has occurred (Pirrotta and Rastelli, 1994). Indeed, it has been shown that high levels of a transactivator can antagonise a PcG protein complex (Zink and Paro, 1995). These hypotheses were recently tested, by analysing the ability of a PRE to establish repression when linked to enhancers with or without HB binding sites (Poux et al., 1996). Particular *Ubx* imaginal disc enhancers were used, which contain no HB binding sites, and direct uniform expression of a reporter gene in imaginal discs, but not in the embryo. When linked to a PRE, repression is established throughout the embryo, regardless of the domains of *hb* or *Ubx* expression, and this repression persists throughout larval development. An early embryonic enhancer (with HB binding sites) combined with a PRE however represses a reporter gene only anterior to parasegment 6. These results favour the idea that the state of activity of the reporter gene during early development is a determining factor for the establishment of a PcG silencing complex.

1.12. Models for Polycomb group protein function

Two distinct mechanisms have been proposed for PcG-mediated silencing. Firstly, analogies with heterochromatin led to a chromatin compaction model, which proposes that complexes of PcG proteins package target loci into heterochromatin-like structures which are inaccessible to *trans*-acting factors (Locke et al., 1988; Paro, 1990). In this scenario, PcG proteins would nucleate at PREs and spread through highly cooperative interactions over entire chromosomal domains.

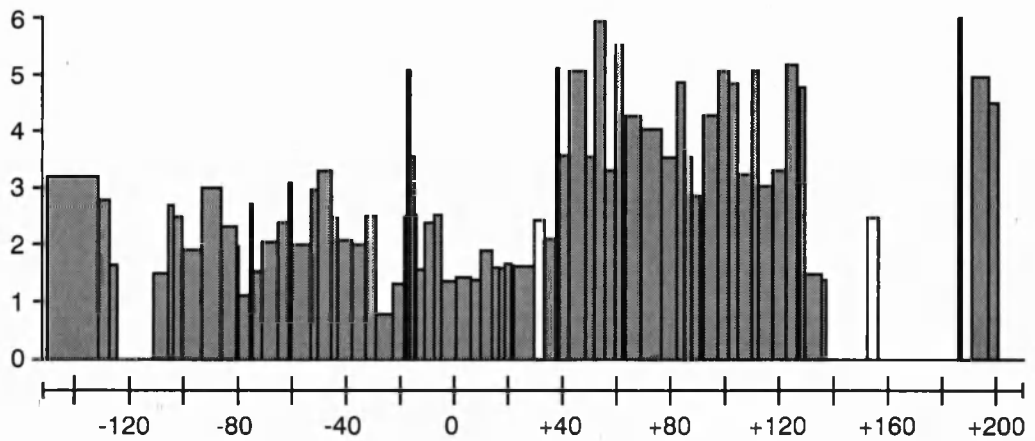
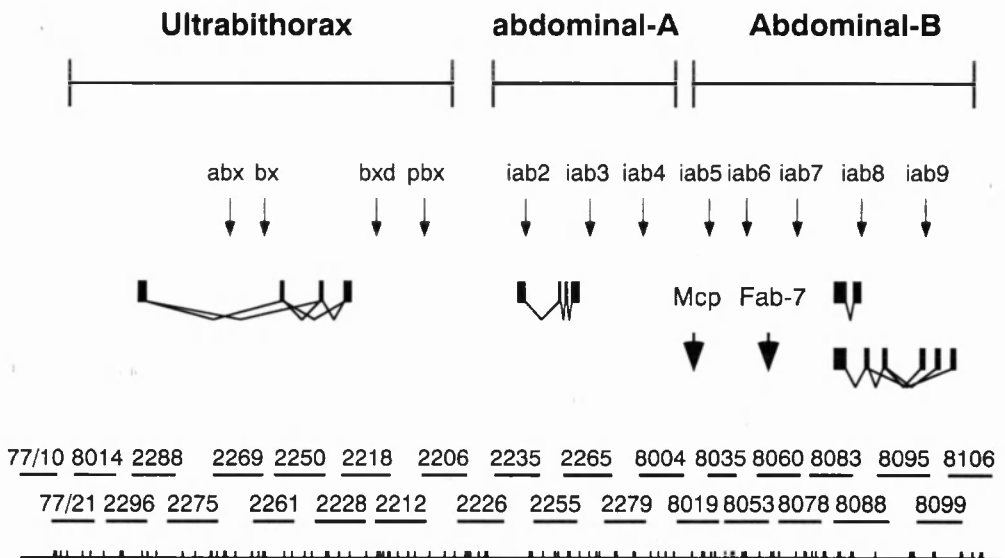
The second model suggests that the binding of PcG proteins to PREs could prevent access of transcription factors, without compaction of target DNA. This could be achieved in several different ways. For example, the presence of PcG protein complexes at PREs could lead to the sequestration of target genes in a particular nuclear compartment, similar to those proposed for heterochromatin or yeast silencing complexes. Compartment boundaries would somehow form a physical barrier for general transcription factors (Paro, 1993). An enhancer interference model on the other hand suggests that PcG proteins bind to a limited number of sites, including PREs, in or near the target gene. These sites may interact together, and interfere with enhancer-promoter looping interactions (Pirrotta and Rastelli, 1994). Alternatively, PcG proteins bound to PREs may directly interfere with promoter function by looping and recruiting the promoter into the complex (Bienz and Müller, 1995). Finally, one consequence of the structures formed by the assembly of PcG complexes may be a reduction in nucleosome mobility, which in itself may decrease transcription factor binding (Kingston et al., 1996).

Experiments have been carried out to test these models of PcG function. The *in vivo* association of PC with target genes was examined by immunoprecipitating formaldehyde cross-linked chromatin from tissue culture cells (Orlando and Paro, 1993). The results indicated that PC is bound extensively throughout the entire regulatory domains of *Ubx* and *abd-A*, which are not expressed in these cells (figure 1.3). This would therefore support a chromatin compaction model for PcG silencing.

A number of other experiments however favour alternative mechanisms of PcG action. Firstly, the accessibility of DNA of the BX-C to restriction endonucleases was examined (Schlosser et al., 1994). No difference in accessibility could be detected either between active and inactive genes in isolated imaginal disc nuclei, or in *Abd-B* DNA from wild type or *Pc* mutant embryos. Secondly, heterologous DNA-binding proteins were used as probes for DNA accessibility in embryos (McCall and Bender, 1996). Binding sites for the yeast transcriptional activator GAL4 or the bacteriophage T7 RNA polymerase were inserted into the *bx* regulatory domain of *Ubx*. Whereas T7 RNA polymerase could direct transcription in all segments of the embryo, GAL4-mediated transcription occurred only in the posterior segments, in which the *bx* region is active. These results argue against simple compaction of DNA which excludes all proteins, but suggest that RNA polymerase II-mediated transcription is impeded, either by compartmentalisation or by interacting with specific molecules required for transcription.

Figure 1.3. Association of Polycomb protein with inactive genes of the bithorax complex.

The organisation of the BX-C is shown at the top of the figure. The intron-exon structure of the *Ubx*, *abd-A* and *Abd-B* transcription units are depicted as black boxes, and the approximate positions of the various regulatory regions are shown by arrows. The line below shows the distribution of EcoRI sites in the BX-C, and the extent of λ clones which cover the region. The lower part of the figure shows the distribution of PC in the complex as deduced from the quantitation of the hybridisation intensities of immunoprecipitated chromatin fragments. The height of the bars represent the absolute values of each individual fragment (in arbitrary units). The open bars indicate the fragments containing repetitive elements (M repeats) that also hybridised with the control fraction without anti-PC antibodies, thus making a precise quantitation impossible. At the bottom of the figure are the coordinates of the BX-C in kilobases, based on the maps of Bender et al. (1983) and Karch et al. (1985). All data from Orlando and Paro (1993).



1.13. Boundary elements in the bithorax complex

The PREs controlling *Abd-B* expression in *iab-5* and *iab-7* are uncovered by the *Mcp* and *Fab-7* deletions respectively (Busturia and Bienz, 1993; Zink and Paro, 1995). These deletions also cause a loss of function of so-called "boundary elements", which are thought to separate the *cis*-regulatory domains of the BX-C (Gyurkovics et al., 1990).

Boundary elements are DNA sequences which are thought to prevent inappropriate enhancer-promoter looping interactions, and therefore to insulate the expression of a gene from the effects of neighbouring sequences (Eissenberg and Elgin, 1991; Schedl and Grosveld, 1995). Apart from *Mcp* and *Fab-7*, two boundary elements of *Drosophila* have been characterised: the "specialised chromatin structure" elements *scs* and *scs'*, which are located flanking two divergently transcribed *heat shock protein 70* (*hsp70*) genes, and the suppressor of Hairy wing (*su(Hw)*) binding region of the gypsy retrotransposon. Whilst the expression of the *w* gene on transposable elements is normally very sensitive to position effects, flanking the *w* gene with *scs* elements results in wild type levels of expression of *w*, regardless of the chromosomal location (Kellum and Schedl, 1991). In addition, the presence of an *scs* element between an enhancer element and a *hsp70-lacZ* reporter gene abolishes activation of the reporter gene by the enhancer element (Kellum and Schedl, 1992). Similarly, insertion of a *gypsy* element into the upstream regulatory region of the *yellow* (*y*) gene causes a loss of *y* expression in certain tissues. Enhancers controlling expression in these tissues were found to be more distal to the promoter than the *gypsy* element (Corces and Geyer, 1991). Therefore two general properties were attributed to boundary elements: they behave as insulators when flanking a gene, and as functional barriers when inserted between an enhancer and a test gene target (Eissenberg and Elgin, 1991). They appear to delimit chromatin domains which correspond to units of autonomous genetic function. These domains do not necessarily correspond to boundaries of structural domains, such as matrix or scaffold attachment regions (section 1.7; discussed in Schedl and Grosveld, 1995; Geyer, 1997).

Two models have been proposed for the mechanism by which domain boundaries function. Firstly, boundaries may assemble nucleoprotein complexes that delimit a chromatin domain, and interactions between regulatory elements are prevented by particular higher-order chromatin structures. Alternatively, a transcriptional decoy model suggests that boundary elements are associated with protein complexes similar to those found at promoters. In this case they inhibit inappropriate enhancer-promoter interactions by intercepting regulatory complexes bound at enhancers (Geyer, 1997).

The *Mcp* and *Fab-7* deletions cause an opposite phenotype to most regulatory mutations of the BX-C, in that they cause posterior transformations of body segments. The *Mcp* deletion lies between *iab-4* and *iab-5*, and causes transformation of the fourth abdominal to the fifth abdominal segment, whereas the *Fab-7* lesion (between *iab-6* and *iab-7*) causes a transformation of the sixth to the seventh abdominal segment (Celniker et al., 1990; Gyurkovics et al., 1990; Karch et al., 1994). The *Fab-7* phenotype was interpreted as resulting from the loss of a boundary element, whose normal function is to block the spreading of the open chromatin configuration of *iab-6* into *iab-7* (Gyurkovics et al., 1990). This hypothesis was supported by the observation that a transgene inserted between *Fab-7* and *iab-7* is activated by *iab-7* enhancers, but not by the more distal *iab-5* and *iab-6* regulatory elements (Galloni et al., 1993). Furthermore, the *Fab-7* and *Mcp* elements are associated with nuclease hypersensitive sites, similar to those observed in the *scs* and *scs'* elements (Galloni et al., 1993; Karch et al., 1994; Udvardy et al., 1985). *Fab-7* was more recently shown to exhibit classical boundary element features such as enhancer blocking activity (Hagstrom et al., 1996; Zhou et al., 1996). However, it should be noted that the normal function of the *Fab-7* boundary is not enhancer blocking, but ensuring the autonomy of the *iab-6* and *iab-7* regulatory domains: the presence of *Fab-7* between *iab-6* and the *Abd-B* promoters does not prevent communication between these elements. In fact, the enhancer blocking activity of *Fab-7* is weaker than that of other boundaries such as *su(Hw)* boundaries (Hagstrom et al., 1996), suggesting that in its normal chromosomal context the *Fab-7* activity is sufficient to prevent communication between *iab-6* and *iab-7*, but not enough to interfere with enhancer-promoter interactions.

In the case of *Fab-7*, the boundary element and the PRE appear to be closely-linked but physically separate elements (Hagstrom et al., 1996; Mihaly et al., 1997), but their functions may be related. In fact, one function of the *Fab-7* boundary may be to stop the spreading of PcG protein complexes into *iab-6* from the *Fab-7* PRE. Whilst the original *Fab-7* deletion (affecting both the PRE and boundary element) clearly showed an invasion of an open chromatin configuration from *iab-6* into *iab-7*, and not vice versa (Gyurkovics et al., 1990), additional *Fab-7* mutations have been recovered in which only the boundary element but not the PRE is deleted (Karch et al., 1994; Mihaly et al., 1997). Interestingly, these deletions apparently do not lead exclusively to the activation of *iab-7* in parasegment 11, but there is competition between ectopic activation of *iab-7* by positive elements in *iab-6* and ectopic silencing of *iab-6* by negative elements in *iab-7* (Mihaly et al., 1997). This supports the idea that one function of the *Fab-7* boundary element may be to limit the spreading of PcG

silencing complexes from the closely-linked PRE. Interestingly, the effects of mutations in the *modifier of mdg4* (*mod(mdg4)*) gene suggest that inhibiting the expansion of chromatin structures may be a more general function of boundary elements: the *mod(mdg4)* gene is essential for the enhancer blocking activity of the gypsy retrotransposon and is also an enhancer of PEV (Gerasimova et al., 1995; Cai and Levine, 1997).

1.14. Genetic and molecular analysis of the trithorax group genes

Most of the *trxG* genes were identified in screens for suppressors of the dominant homeotic phenotypes of *Pc* or *Antp* (Kennison and Tamkun, 1988). Mutations in any of these genes mimic loss of function mutations of the BX-C and ANT-C genes, and therefore they were considered to be activators of the homeotic genes. In fact, they appear to counteract the formation of repressive PcG chromatin structures, and to maintain their target genes in an open configuration which is accessible to transcription factors.

The *trx* gene is the most genetically well-characterised member of the *trx-G*. Mutations in *trx* suppress the homeotic transformations caused by *Pc* or *esc* mutations (Ingham, 1983; Kennison and Tamkun, 1988). Mutant embryos and adults show anterior transformations of segments (Ingham and Whittle, 1980; Ingham, 1983; Ingham, 1985), and homozygous embryos express reduced levels of a variety of homeotic genes, consistent with the specific transformations seen in embryonic cuticles (Breen and Harte, 1993). In addition *trx* is necessary for the maintenance of *en* expression (Breen et al., 1995). Similar results were obtained with mutations in a number of other *trx-G* genes (Tamkun et al., 1992; Farkas et al., 1994; LaJeunesse and Shearn, 1995). Interestingly, mutations in the *absent*, *small*, and *homeotic discs 1 and 2* (*ash 1 and 2*) genes showed different alterations of homeotic gene expression, suggesting that all *trxG* proteins do not function together in a complex (LaJeunesse and Shearn, 1995). Indeed, the *trxG* genes appear to be very heterogeneous, which may be expected as there are many potential steps which they may regulate in the expression of homeotic genes.

Five *trxG* genes have been characterised molecularly (table 1.2). Both *trx* and *ash1* contain a SET domain which is also seen in the mouse *trx* homologue *Mll* and the PcG gene *E(z)* (Mazo et al., 1990; Tripoulas et al., 1996; Yu et al., 1995). Furthermore, they both bind to partially overlapping sites in polytene chromosomes (Kuzin et al., 1994; Chinwalla et al., 1995; Tripoulas et al., 1996). In the case of *trx*, these sites include the positions of the BX-C, ANT-C and *en*, suggesting that they act as direct transcriptional activators (Chinwalla et al., 1995).

Table 1.2. The cloned trithorax group genes

Gene	Cytological location	Conserved motifs	References
<i>trithorax (trx)</i>	88B	SET domain PHD finger	Mazo et al., 1990
<i>brahma (brm)</i>	72AB	ATPase motif bromodomain	Tamkun et al., 1992
<i>trithorax-like (Trl)</i>	70F	zinc finger BTB domain	Farkas et al., 1994
<i>absent, small or homeotic discs 1 (ash1)</i>	76B	SET domain PHD finger	Tripoulas et al., 1996
<i>absent, small or homeotic discs 2 (ash2)</i>	96A	PHD finger	Adamson and Shearn, 1996

Cloning of the *brahma (brm)* and *Trithorax-like (Trl)* genes led to the discovery of some interesting homologies, which shed light on the function of these genes. The *brm* gene product was shown to be the homologue of the yeast SWI2/SNF2 (yeast mating type switching/sucrose non-fermenting) transcriptional activator (Tamkun et al., 1992), with which it shares two highly conserved regions (the bromodomain and the ATPase motif). Secondly, the *Trl* gene was shown to encode GAGA factor, a transcriptional activator first characterised as binding to GA/CT-rich target sequences in the *Ubx* promoter (Biggin and Tjian, 1988; Farkas et al., 1994). Both SWI2/SNF2 and GAGA factor appear to function not as transcription factors, but by remodelling chromatin structures.

1.15. Chromatin remodelling activities in *Drosophila*

The yeast SWI/SNF proteins, including SWI2/SNF2, are members of a large, 11 subunit multimeric complex (Cairns et al., 1994; Cote et al., 1994; Peterson et al., 1994; Treich et al., 1995). The SWI/SNF genes were first identified as being necessary for the normal transcription of the *HO* (encoding an endonuclease required for mating type switching) or *SUC2* genes (Winston and Carlson, 1992; Kingston et al., 1996). Genetic studies have indicated that the SWI/SNF complex disrupts chromatin structure. Firstly, point mutations in either histone H3 or H4 have a SIN (switch-independent) phenotype: that is, they partially

suppress the requirement for the SWI genes in transcriptional activation (Kruger et al., 1995). Another SIN gene, SIN1, encodes a protein with homology to the high mobility group protein 1 (HMG-1), and is thus also likely to be a structural component of chromatin (Kruger and Herskowitz, 1991). Finally, reducing the amount of histones H2A and H2B in the cell suppresses *swi/snf* phenotypes (Hirschhorn et al., 1992). These results therefore suggest that destabilisation of chromatin structure can reduce the requirement for SWI/SNF activity.

The SWI/SNF complex has ATPase activity (Laurent et al., 1993), and is thought to function by disrupting chromatin structure at promoters, and thus increasing access to the transcriptional machinery. Indeed, chromatin structure at the SUC2 promoter is altered in *snf2* and *snf5* mutants (Hirschhorn et al., 1992), and a purified SWI/SNF complex can modify nucleosomal structure in an ATP-dependent manner (Cote et al., 1994). This nucleosomal disruption facilitates the *in vitro* binding of transcription factors to a nucleosomal template. The mechanism does not appear to involve displacement of histones from DNA, but histone-DNA contacts are modified, probably by altering the degree of supercoiling of nucleosomal DNA (Kwon et al., 1994; Wilson et al., 1996).

The recently identified yeast RSC (remodel the structure of chromatin) complex has similar properties to the SWI/SNF complex, and several of their components (including SWI2/SNF2) are structurally related (Cairns et al., 1996). Moreover, RSC is 10-fold more abundant than SWI/SNF, suggesting that it may have a more general role in transcriptional activation. The SWI/SNF complex has been reported to be associated in stoichiometric amounts with the RNA polymerase II holoenzyme, which would allow the specific targeting of SWI/SNF activity to promoter regions (Wilson et al., 1996). This result is however controversial (see Cairns et al., 1996), and would not explain the specificity of SWI/SNF for only a subset of RNA polymerase II-transcribed genes.

The function of the SWI/SNF complex appears to be conserved in higher eukaryotes. Two mammalian homologues of SWI2/SNF2 have been identified (Khavari et al., 1993; Muchardt and Yaniv, 1993), which are components of distinct multimeric complexes that may function in different cell types (Wang et al., 1996). Purified human SWI/SNF complexes have the ability to disrupt nucleosomes and to enhance the *in vitro* binding of transcription factors and TATA-binding protein to nucleosomal DNA (Kwon et al., 1994; Imbalzano et al., 1994; Wang et al., 1996). Similarly, the *Drosophila* *brm* and *snf5-related 1* (*snr1*) gene products are members of a multimeric protein complex (Dingwall et al., 1995). Interestingly, mutations in *snr1* interact with *trx*, enhancing the anterior transformation of the fifth abdominal segment observed in *trx* heterozygotes (Dingwall et al., 1995). This suggests that

snr1, and presumably other members of the *Drosophila* SWI/SNF complex, may be classed as *trxG* genes. By analogy to their mammalian and yeast homologues, the BRM/SNR1 complex was suggested to function by counteracting the repressive effects of PcG proteins and maintaining a chromatin structure which allows activator binding.

GAGA factor is also involved in organising accessible chromatin structures, and its function has been well-studied at the promoters of the *Drosophila hsp26* and *hsp70* genes. These are "preset" promoters, characterised by the presence of DNaseI hypersensitive sites in the TATA box and upstream regulatory elements (heat shock elements, HSEs). Transcription factor II D (TFIID) and a paused RNA polymerase II molecule are also constitutively bound to the promoter. This promoter structure keeps the HSEs accessible, and enables the rapid binding of heat shock factor (HSF) after heat shock or other environmental stress. HSF then activates transcription by releasing the paused polymerase (Wallrath et al., 1994). The binding of GAGA factor to GA/CT elements in the promoter is essential for the formation of the preset chromatin structure (Lu et al., 1993a). In fact, GAGA factor is constitutively associated with the promoter regions of both *hsp26* and *hsp70*. After heat shock, GAGA factor is also recruited to the transcription units, where it may allow efficient elongation by RNA polymerase II (O'Brien et al., 1995). It has been suggested that GAGA factor may function by counteracting repression mediated by the linker histone H1, and thus allowing mobility of nucleosomes (Croston et al., 1991; Wall et al., 1995).

An *in vitro* assay for nucleosome assembly was developed, based on the ability of *Drosophila* extracts to assemble long arrays of regularly-spaced nucleosomes on plasmid DNA templates (Becker and Wu, 1992). If GAGA factor is added either before or after nucleosome assembly, it is able to disrupt nucleosome structure at the promoter region of *hsp70*: the promoter region becomes sensitive to DNaseI digestion and the nucleosomes are redistributed (Tsukiyama et al., 1994). Interestingly, the ability of GAGA factor to disrupt preassembled nucleosomes is reduced in chromatin containing histone H1. Furthermore, the crystal structure of nucleosome-associated GAGA factor suggests that binding of GAGA factor and histone H1 to linker DNA may be mutually exclusive events, thus confirming earlier observations that GAGA factor can relieve histone H1 silencing (Omichinski et al., 1997; Croston et al., 1991).

GAGA factor-dependent nucleosome alterations are facilitated by a nucleosome remodelling factor (NURF) (Tsukiyama and Wu, 1995). NURF is thought to alter nucleosome structure in an ATP-dependent manner, and thus enable GAGA factor binding to GA/CT repeats in the promoter; the binding of GAGA factor then maintains an "open" chromatin

configuration. NURF is composed of four major subunits, and is distinct from the SWI/SNF complex. In addition, its mode of activity is different to that of SWI/SNF, as its ATPase activity is stimulated by nucleosomes rather than by free DNA. Interestingly, one component of NURF was identified as the imitation switch (ISWI) protein, which is highly related to SWI2/SNF2 in its ATPase domain (Tsukiyama et al., 1995). Therefore, a battery of SWI2-related proteins appear to exist in the cell, which are members of different complexes (SWI/SNF and RSC in yeast; BRM/SNR1 and NURF in *Drosophila*) and which may have differing specificities for remodelling DNA.

1.16. The interaction between Polycomb and trithorax group gene products

The antagonistic genetic interactions between the *trxG* and the PcG genes suggest that the *trxG* gene products function to counteract PcG-mediated silencing. The exact mechanism by which this occurs is still obscure, and the heterogeneity of the *trxG* gene products may in fact be indicative of multiple modes of *trxG* action. Firstly, the *trxG* genes may have direct antirepression effects on the PcG silencing mechanism, for example they may interact directly with members of the PcG protein complex to disturb the establishment or the stability of the complex. Alternatively, the effects of the *trxG* genes may be more indirect: by acting as direct transcriptional activators or by allowing other activators to bind, PcG silencing may be impeded (Paro and Harte, 1996). For example the binding of GAL4, a strong transcriptional activator, can counteract PcG silencing, even after the assembly of PcG protein complexes (Zink and Paro, 1995).

The study of PC and TRX has shown that the antagonistic genetic interactions between *trx* and the PcG genes do not appear to be a result of competition for exclusive binding to target genes. Many of the TRX binding sites on the polytene chromosomes of salivary glands coincide with those of PC (Chinwalla et al., 1995). Moreover, PC and TRX bind simultaneously to a 14 kb fragment containing the *bxd* PRE in a reporter construct (Chan et al., 1994; Chinwalla et al., 1995). This suggests that trithorax response elements (TREs) and PREs are closely-linked, although it is not clear whether the situation at the *bxd* PRE is representative of all PREs/TREs, or if it is an isolated case. The fact that *Ubx* is not expressed in salivary glands indicates that TRX binds constitutively to its target genes, irrespective of the expression status.

TRX is unlikely to act as a conventional transcriptional activator, as its binding site in the *bxd* PRE has no enhancer activity in embryos. On the other hand in transient expression assays in *Drosophila* tissue culture cells, TRX and PC can function through a 440 bp element

of this PRE to activate or repress expression from the *Ubx* promoter respectively (Chang et al., 1995). However, only 3 to 4-fold activation of transcription is observed with TRX, again suggesting that it is itself not a strong activator. TRX-dependent activity can be further reduced by increasing the amount of PC expressed in these cells, suggesting that PC and TRX interact competitively (Chang et al., 1995). It is not clear if this reflects a direct physical interaction between the members of the *trxG* and *PcG*. However the presence of a common sequence motif, the SET domain, in TRX, ASH1 and E(Z) suggests that they may interact with a common molecular target (Mazo et al., 1990; Tripoulas et al., 1996; Jones and Gelbart, 1993).

Ultimately, an understanding of the manner in which *PcG* and *trxG* proteins interact may require better characterisation of TRX target elements. In addition, it is not yet known if other *trxG* proteins act at the same sites as TRX: GAGA factor for example is best characterised as being associated with promoter elements. The phenotype of heterozygous *Trl* adults however, in which the sixth abdominal segment is most conspicuously transformed, suggests that GAGA factor is also involved in maintaining an active chromatin configuration in distant regulatory elements such as *iab-6* (Farkas et al., 1994). By analogy to its function in the *hsp* promoters, GAGA factor may bind to target elements in a particular parasegmental regulatory region to maintain an "open" chromatin configuration in those domains in which a particular homeotic gene is active. BRM/SNR1 and NURF on the other hand may function more transiently to open chromatin and allow the access of DNA-binding factors such as TRX or GAGA.

1.17. Aims of the thesis

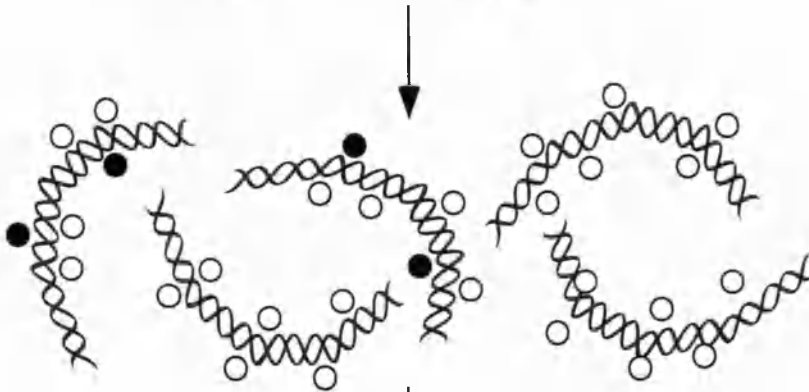
The experiments described in this thesis were designed to further knowledge about the interactions of the *PcG* and *trxG* proteins with their target genes, in particular the *BX-C* and *en*, by immunoprecipitating these proteins from *in vivo* cross-linked chromatin. A schematic diagram of the cross-linking and immunoprecipitation technique is shown in figure 1.4. *Drosophila* Schneider cells are cross-linked with formaldehyde, and the chromatin is purified. Specific antibodies are then used to immunoprecipitate DNA which is covalently cross-linked to the protein of interest. Formaldehyde cross-links are reversed, and the DNA purified (Solomon et al., 1988; Orlando and Paro, 1993; Orlando et al., 1997). This procedure allows the isolation of a small quantity of DNA, which is enriched for the specific protein-associated elements. If potential target sequences are available, they can be used to probe a slot blot containing immunoprecipitated DNA, and to determine enrichments over control

Figure 1.4. Scheme of the formaldehyde cross-linking and chromatin immunoprecipitation method.

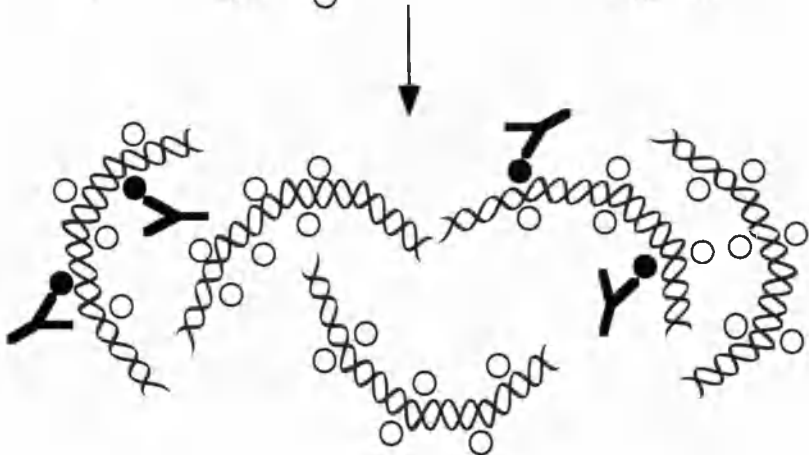
Drosophila Schneider cells are cross-linked with formaldehyde, before sonicating to produce soluble chromatin of an average size of 1 kb, in which proteins (circles) are covalently cross-linked to DNA. Chromatin is purified on a caesium chloride gradient, before incubating with antibodies which recognise a particular DNA-binding or chromatin-associated protein (black circles). Immunocomplexes are purified on Protein-A Sepharose, the cross-links are reversed and the co-immunoprecipitated DNA is purified. Approximately 1 ng DNA is isolated, and this DNA is subjected to linker-modified PCR to generate sufficient DNA to be used as a probe on a Southern. Hybridisation of the immunoprecipitated probe to each restriction fragment of a genomic walk is then quantitatively analysed.



Cross-link whole cells
with formaldehyde



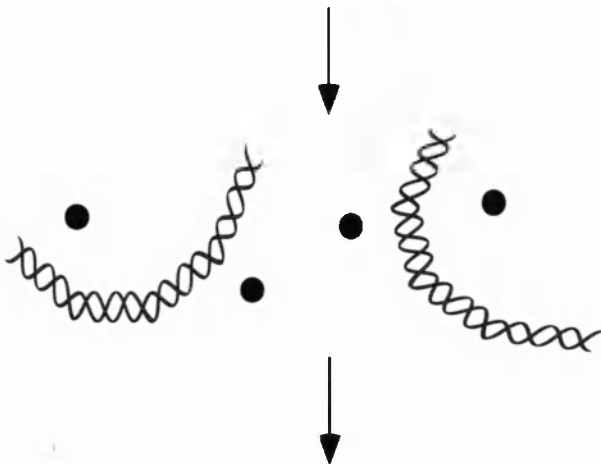
Sonicate to produce
sheared, soluble
chromatin



Purify on caesium
chloride gradient

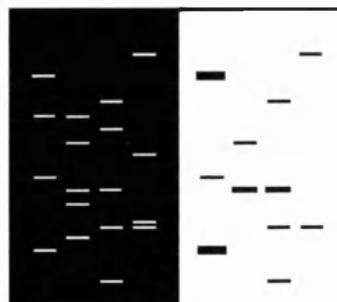
Immunoprecipitate
with specific
antibodies

Purify immuno-
complexes on
Protein-A Sepharose



Reverse cross-links
and purify DNA

Agarose gel Southern



Amplify by ligation-
mediated PCR and
use as probe on
Southern of genomic
region of interest

immunoprecipitations without antibody. Furthermore, a modification of the procedure utilises linker-modified PCR to amplify the immunoprecipitated DNA, which can then be used as a probe on a Southern of large genomic walks (Orlando and Paro, 1993; Orlando et al., 1997). A relatively short exposure to formaldehyde (8 minutes) was sufficient to detect DNA-histone interactions (Solomon et al., 1988). However, the analysis of non DNA-binding, chromatin-associated factors could be achieved by utilising longer cross-linking times (Orlando and Paro, 1993; Zhao et al., 1995).

Previous work from this laboratory utilised this approach to analyse the distribution of PC on the entire BX-C in tissue culture cells (Orlando and Paro, 1993). In chapter 2 an improvement to the formaldehyde cross-linking technique is described, which eliminates PCR artifacts associated with the original method and allows high resolution *in vivo* mapping of proteins on their target genes.

The work in this thesis addresses three main points concerning the function of the PcG and trxG proteins. Firstly, mapping the distribution of PC at high resolution on the BX-C reveals that the majority of PC is associated with PREs: this result has important implications for understanding the mechanisms by which PcG proteins silence their target genes (chapter 3).

Secondly, does GAGA factor have a function at distant regulatory elements as well as at promoters (see section 1.16)? The mutant phenotype of adult *Trl* heterozygotes suggests that this may be the case (Farkas et al., 1994), but no molecular data was so far available. In fact, the data obtained from mapping GAGA factor binding sites in the BX-C (chapter 3) indicates that GAGA factor may have multiple roles in upstream regulatory domains, particularly at PREs and at the *Fab-7* boundary element.

Finally, how is the fact that PcG proteins are members of a common multimeric complex reconciled with the finding that PcG mutants have different phenotypes (section 1.5)? The production of antibodies against PH and PSC is shown in chapter 4, together with evidence that PH and PSC form a complex with PC. The distribution of PH and PSC on the *en-inv* locus and the *Abd-B* gene suggests that the composition of the PcG complex varies at different loci (chapter 5): the functions of these multiple complexes may contribute to the differences in PcG mutant phenotypes.

- CHAPTER 2 -

Improvement of an *in vivo* formaldehyde cross-linking method for the analysis of protein-DNA interactions

2.1. Introduction

A number of methods have been described for the analysis of protein-DNA interactions *in vivo*. Two general approaches have been described: either cross-linking or non-cross-linking methods. An example of a non-cross-linking approach involved the preparation of soluble chromatin by subjecting endonuclease-treated nuclei to a hypotonic shock (Gould et al., 1990). The use of specific antibodies to immunoprecipitate from this chromatin allowed the identification of a number of genes which are potential target genes of *Ubx* (Gould and White, 1992; Strutt and White, 1994).

Cross-linking methods have been more widely used for probing protein-DNA interactions. For example, an *in vivo* UV cross-linking assay was used to locate RNA polymerase II in the heat shock genes of *Drosophila* (Gilmour and Lis, 1986). UV light induces cross-linking only between those protein and nucleic acid residues in intimate contact with each other, and the cross-linking is proportional to the occupancy of a protein on its binding site. Protein-DNA complexes could be isolated by immunoprecipitation, and the DNA characterised by Southern blotting. The sensitivity of this method was subsequently improved to allow the mapping of the *in vivo* binding sites of sequence-specific transcription factors (Walter et al., 1994; O'Brien et al., 1995).

An alternative *in vivo* cross-linking approach was also developed, which used formaldehyde to introduce protein-DNA cross-links (see figure 1.4). This method was used to analyse the chromatin structure of the SV40 origin of replication and to assess alterations in nucleosomal structures in the *hsp70* promoter after heat shock (Solomon and Varshavsky, 1985; Solomon et al., 1988). Whereas UV and other cross-linking agents damage DNA extensively upon prolonged exposure, formaldehyde does not react with free double-stranded DNA. However, protein-DNA, protein-RNA and protein-protein cross-links are induced very rapidly upon formaldehyde treatment, thus creating a stable structure which is believed to prevent redistribution of cellular components. Furthermore, cross-links are reversible, allowing purification of DNA for further analysis (Solomon and Varshavsky, 1985).

Using this methodology, the *Mcp* PRE was shown to be highly enriched in chromatin immunoprecipitated with PC antibodies (Orlando and Paro, 1993). In addition, PC immunoprecipitated DNA was amplified by linker-modified PCR and used as a probe on a Southern of the entire BX-C walk. These results indicated that PC was not associated with the *Abd-B* gene, which is expressed in Schneider cells. However, PC was associated with extended regulatory regions of the inactive genes *Ubx* and *abd-A* (figure 1.3; Orlando and Paro, 1993). As all DNA in the inactive domain was associated with PC, and not only isolated regulatory elements, this result was consistent with a compaction model for PcG function.

2.2. Comparison of the hybridisation of control and Polycomb-immunoprecipitated DNA to the bithorax complex walk

The hybridisation of PC-immunoprecipitated DNA to a Southern of the BX-C walk revealed an irregular distribution on the inactive genes (Orlando and Paro, 1993). Whilst PC appeared to cover the entire inactive domain, a number of distinct peaks were also observed. It was suggested that these peaks correspond to important regulatory elements of the bithorax complex; for example, the *Fab-7* PRE and sequences close to the *Mcp* PRE were strongly enriched by PC immunoprecipitations.

These experiments were repeated as previously described, with additional controls. DNA from P1 bacteriophage clones covering the BX-C was used for Southern blots. The complete sequence of the bithorax complex is now available (Martin et al., 1995), and thus individual restriction fragments can be unambiguously aligned with their position on the walk. Due to the large size (80 - 100 kb) of insert DNA contained within P1 vectors, many doublet bands were produced after restriction digestion and gel electrophoresis (figure 2.1a). Therefore each clone was subjected to two separate digestions, with *EcoRI* or *SalI/XhoI*, to allow analysis of all regions of the walk.

PCR-amplified PC immunoprecipitated DNA hybridises strongly to most fragments in the inactive region of the BX-C (represented in the 5 P1 clones, figure 2.1b), but not to λ clones covering the *Abd-B* gene, as previously reported (Orlando and Paro, 1993). Control immunoprecipitations without antibody also yield DNA, presumed to be nonspecifically purified. This DNA (after PCR amplification) hybridises to the BX-C, and can be seen by a long exposure of the Southern to a Phosphorimager screen (figure 2.1c). This would be expected, as an equivalent amount of *Drosophila* genomic DNA also gives a detectable signal (see figure 2.4c).

Figure 2.1. Southern hybridisation of immunoprecipitated, PCR-amplified DNA to the bithorax complex genomic walk.

(a) 1 µg DNA from λ bacteriophage, or 2 µg DNA from P1 bacteriophage was digested with restriction enzymes and separated on a 0.6 % agarose gel. The λ clones cover the distal region of the BX-C; lane 1 (λ8106), lane 2 (λ8099), lane 3 (λ8095), lane 4 (λ8088), lane 5 (λ8083). Lanes 1, 2 and 3 were digested with EcoRI and HindIII, and lanes 4 and 5 with EcoRI. P1 clones cover the remaining 290 kb of the BX-C, and each clone was subjected to two separate digestions, with EcoRI or SalI/XhoI. Distal clones are to the left of the gel, proximal clones to the right; lane 1 (DS04698), lane 2 (DS00846), lane 3 (DS03408), lane 4 (DS03126), lane 5 (DS05563). Lane M is a molecular weight marker, consisting of λ DNA digested with EcoRI/HindIII; the approximate positions of 5.0 kb, 4.2 kb, 2.0 kb and 1.0 kb are indicated adjacent to the marker.

(b) Hybridisation of DNA from a PC immunoprecipitation to a Southern filter of the gel shown in (a).

(c) An identical filter to that in (b) hybridised to DNA purified from control immunoprecipitations, without antibody. Note that the overall signal is weaker, but that the hybridisation pattern resembles that in (b). This is particularly evident for the bracketted restriction fragments of λ8106 and λ8099 (λ clone lanes 1 and 2).

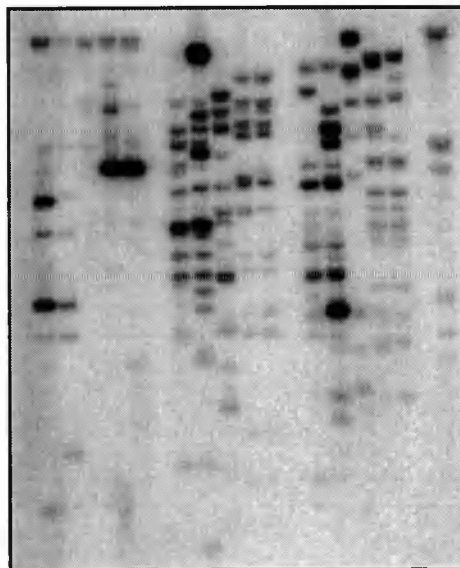
(d) An identical filter probed with *Drosophila* genomic DNA which was amplified by restriction enzyme-mediated linker-modified PCR. Again, note the similarity to (b) and (c).

a

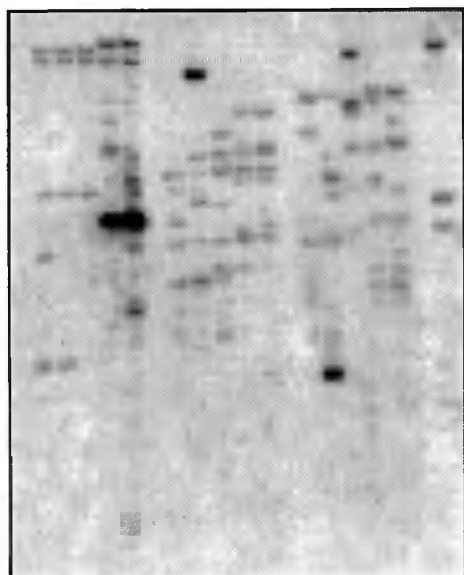
λ clones P1 clones P1 clones
 EcoR1 Sal1/Xho1
1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 M

**b****PC IP**

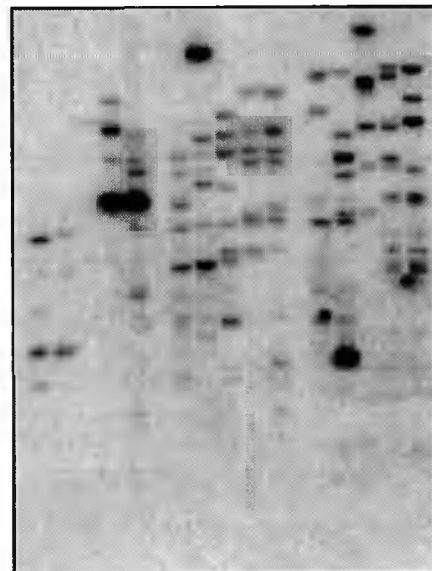
λ clones P1 clones P1 clones
 EcoR1 Sal1/Xho1
1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 M

**c****Control IP**

λ clones P1 clones P1 clones
 EcoR1 Sal1/Xho1
1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 M

**d****PCR amplified
genomic DNA**

λ clones P1 clones P1 clones
 EcoR1 Sal1/Xho1
1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 M



Whilst PC immunoprecipitations clearly enrich for DNA in the *Ubx* and *abd-A* regions of the BX-C, as expected, surprisingly the overall pattern of hybridisation is in fact similar in both control and PC-immunoprecipitations. In particular, many DNA fragments covering the inactive *Abd-B* domain hybridise to neither PC nor control immunoprecipitation probes (compare λ clone lanes 3 and 4, figure 2.1b and c). Furthermore, fragments which hybridise most strongly to the PC immunoprecipitated probe also hybridise most strongly to the control immunoprecipitated probe (compare the 5 bracketed fragments of λ clone 1, figure 2.1b and c). Therefore, this strong hybridisation is likely to be due to the methodology and does not necessarily represent strong PC binding sites.

Hybridisation of PC and control immunoprecipitated DNA probes to EcoRI fragments was quantitated. The resulting values were normalised to account for molecular weight differences, and plotted according to position on the BX-C walk (figure 2.2a). To emphasize the overlap between the hybridisation profiles, both quantitations were plotted on the same graph; PC hybridisation is shown as grey bars and control hybridisation as white bars. The pattern of hybridisation of PC-immunoprecipitated DNA is largely identical to that seen by Orlando and Paro (1993), with a number of small differences caused by incorrect mapping of the location of restriction fragments in the earlier study. However, the *Fab-7* element is poorly enriched by PC immunoprecipitations in the present experiments, in contrast to the strong enrichment previously reported.

2.3. Effect of NdeII site distribution on PCR amplification efficiency

The similarity in the hybridisation profiles of control and PC immunoprecipitated DNA could be due either to an inherent problem with the immunoprecipitation itself (i.e. certain DNA sequences are selectively enriched in control immunoprecipitations), or to uneven PCR amplification of the immunoprecipitated DNA. To test the second possibility, *Drosophila* genomic DNA was amplified using linker-modified PCR, in the same way as for the immunoprecipitated DNA samples, and hybridised to the BX-C (figure 2.1d). Indeed, the hybridisation is strikingly similar to that seen by control immunoprecipitations, thus confirming that the PCR amplification strategy is a major cause of the uneven hybridisation profiles produced by immunoprecipitated DNA probes.

Immunoprecipitated DNA was amplified using a linker-modified PCR strategy, in which purified DNA is digested with NdeII, a four-base cutter restriction endonuclease. Appropriate linkers are then attached to the restricted ends, and used as sites for annealing PCR primers. NdeII digestion results in DNA fragments of an average size of 500 bp;

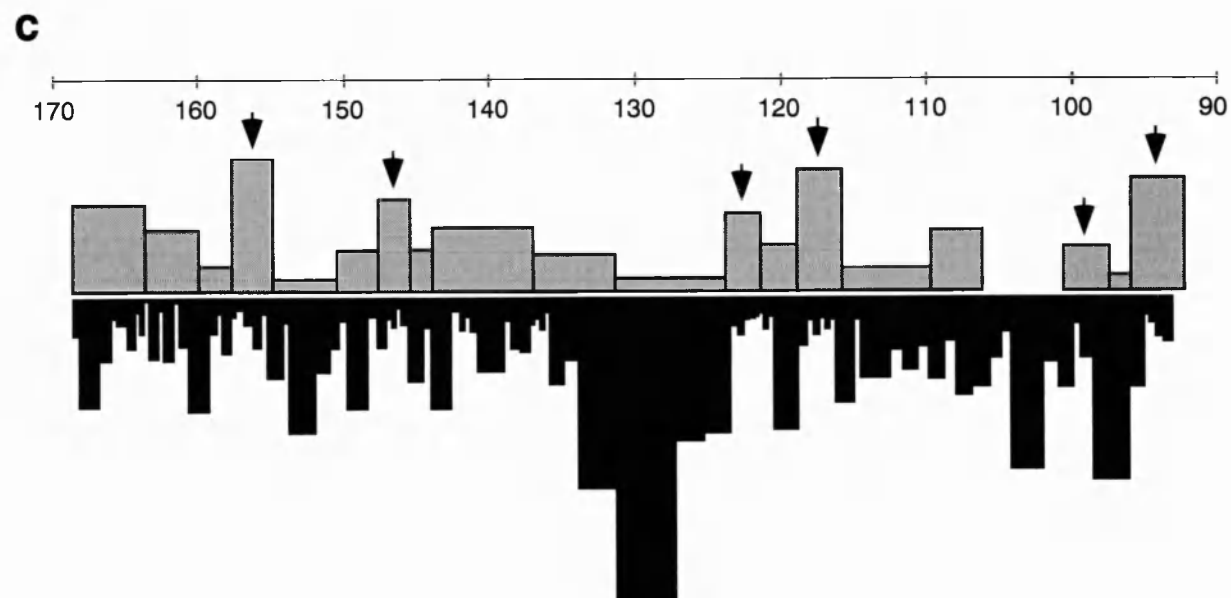
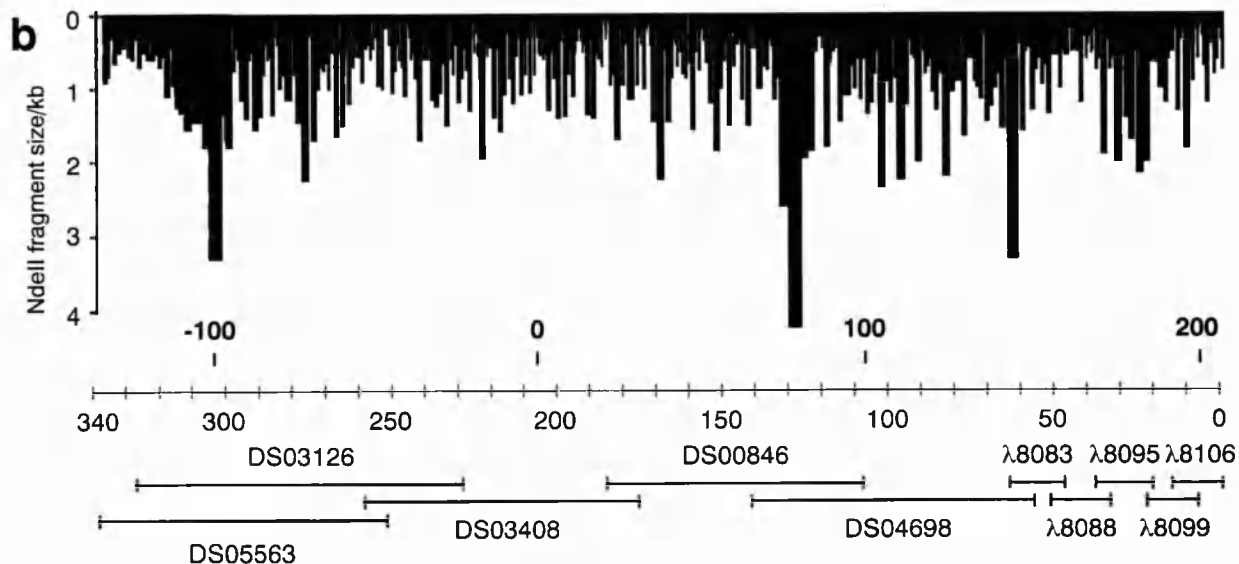
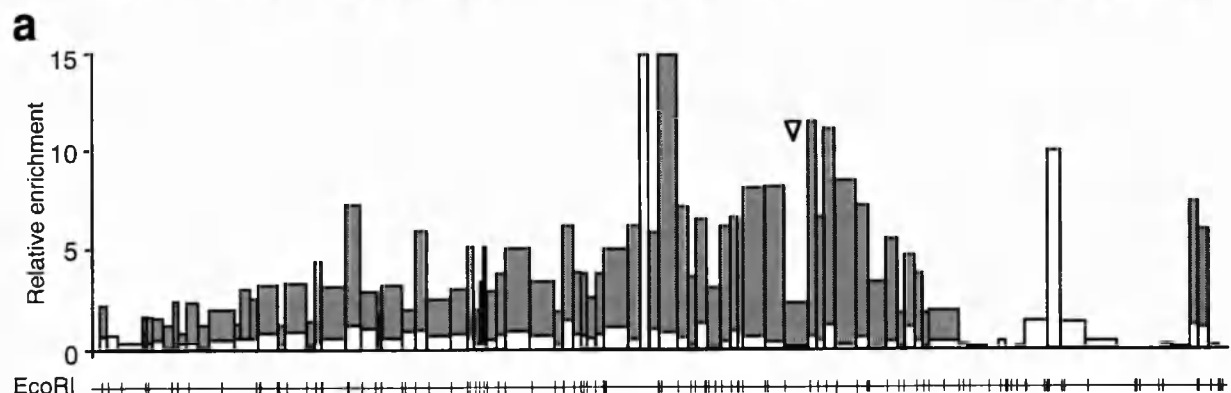
Figure 2.2. Amplification of immunoprecipitated DNA is dependent on NdeII site distribution

The BX-C is depicted with proximal to the left and distal to the right. The various regulatory regions of the BX-C (*abx* to *iab-9*) are indicated at the top of the figure. The exon structure of the three homeotic genes is shown in black, and that of other transcripts/ORFs in grey (AHCY, S-adenosyl homocysteine hydrolase-like; GLU, glucose transporter-like; W, low density lipoprotein (LDL) receptor-like repeats; X, serine protease-like; Y, chaperonin-containing t-complex protein-1 γ subunit-like; Z, no-on transient A-like; data from Martin et al., 1995). Also marked is the position of the *Fab-7* deletion. At the bottom is shown the extent of the P1 or λ clones used in the analysis. The traditional map coordinates are shown in bold type (Bender et al., 1983; Karch et al., 1985), and in normal type are the coordinates based on the complete sequence of the BX-C (0 - 340) (Martin et al., 1995). In the middle is shown the restriction enzyme sites for EcoRI (E); more detailed restriction site information is in Appendix A1.

(a) The hybridisation signals on the Southern filters in figure 2.1b and 2.1c were quantitated, and the resulting values were normalised to account for molecular weight differences and plotted on the map of the BX-C. Hybridisation of PC immunoprecipitated DNA is depicted as grey bars, and that of control immunoprecipitated DNA as white bars (repetitive elements give very strong hybridisation, even in the control). Data was taken from at least two immunoprecipitation experiments, carried out on independent cross-linked chromatin preparations. The scale bar to the left indicates relative strength of hybridisation signal, and is in arbitrary units. The arrowhead indicates a 7.6 kb EcoRI fragment which is poorly amplified due to the wide spacing of NdeII sites (panel (b)).

(b) NdeII site distribution was plotted on the bithorax complex walk as a function of fragment size. The vertical scale bar to the left is fragment size in kb.

(c) An enlargement of the profiles in panels (a) and (b), in the region between coordinates 90 and 170. The upper profile shows relative enrichments of EcoRI fragments in control immunoprecipitations, and the lower profile is plotted as a function of NdeII fragment size. The arrows point to several EcoRI fragments, which are well represented in PCR-amplified control immunoprecipitated DNA, and which contain NdeII fragments of a small average size.



however fragment size varies widely, with many above 1.5 kb (see figure 2.2b). In fact, this uneven distribution of NdeII sites is the probable cause of the unequal amplification of immunoprecipitated DNA, as small DNA fragments amplify much more efficiently than large fragments (Lüdecke et al., 1989). Furthermore, for primer extension stages of PCR amplification 1 minute per kb is usually considered appropriate; the current PCR strategy utilises only 1 minute, and thus it is likely that DNA molecules longer than about 1 kb have insufficient time to extend to completion.

The NdeII site distribution was plotted as a function of fragment size on the BX-C walk (figure 2.2b). A number of EcoRI fragments which hybridise poorly to PC and control immunoprecipitated probes contain large NdeII fragments; most clearly a 7.6 kb EcoRI fragment (arrowhead in figure 2.2a) which contains NdeII fragments of size 1.8 kb, 1.9 kb and 4.2 kb. Elsewhere in the *abd-A* region of the BX-C the apparent peaks and troughs seen in PC immunoprecipitated DNA hybridisations correlate with regions of small and large NdeII fragments (figure 2.2c). In other regions of the BX-C this effect is less dramatic, presumably as most EcoRI fragments contain a more average NdeII site distribution. In any case, it is apparent that the resolution of mapping of PC binding in the BX-C is biased by the NdeII digestion step of the linker-modified PCR strategy.

2.4. Blunt-end linker-modified PCR

An attempt was made to devise a new, simpler PCR method, in which blunt-ended linkers were attached directly to the sonicated DNA fragments generated by the immunoprecipitation. Such DNA fragments would be expected to have a random distribution with respect to a genomic walk, and thus all regions of the walk should amplify approximately linearly. However, sonicated DNA fragments may not be a good substrate for ligation of a blunt linker, as sonication is likely to shear DNA molecules such that the resulting fragments are blunt, and with 5' or 3' overhangs, and phosphorylated or non-phosphorylated. Indeed, cloning of sonicated DNA fragments for other purposes, for example shot-gun sequencing, is reported to occur at very low efficiency for this reason. Therefore, as the DNA from immunoprecipitation experiments is limiting (typically each immunoprecipitation experiment yields 1 ng of DNA), this could be a potential problem.

Plasmid or *Drosophila* genomic DNA was either digested with blunt end restriction endonucleases or sonicated to produce DNA fragments mimicking those resulting from immunoprecipitation experiments. This DNA was ligated to different concentrations of blunt-ended linker molecules, and PCR amplification was carried out, before loading a

Figure 2.3. Amplification of DNA using blunt-end linker-modified PCR

(a) 1 ng RsaI/DraI digested Bluescript (pBS, lanes 1 - 3), 1 ng RsaI digested *Drosophila* genomic DNA (Dm DNA, lanes 4 - 6) or 1 ng sonicated *Drosophila* genomic DNA (lanes 7 - 9) was ligated to a blunt linker and PCR amplified. 1/20th of the reaction product was then separated on a 1.2 % agarose gel. Linker concentration was varied; lanes, 1, 4 and 7, 0.1 μ M linker; lanes 2, 5 and 8, 0.5 μ M linker; lanes 3, 6, and 9, 2.5 μ M linker. Lane M is a molecular weight marker (Boehringer), consisting of a mixture of BglI digested pBR328 and HinfI digested pBR328.

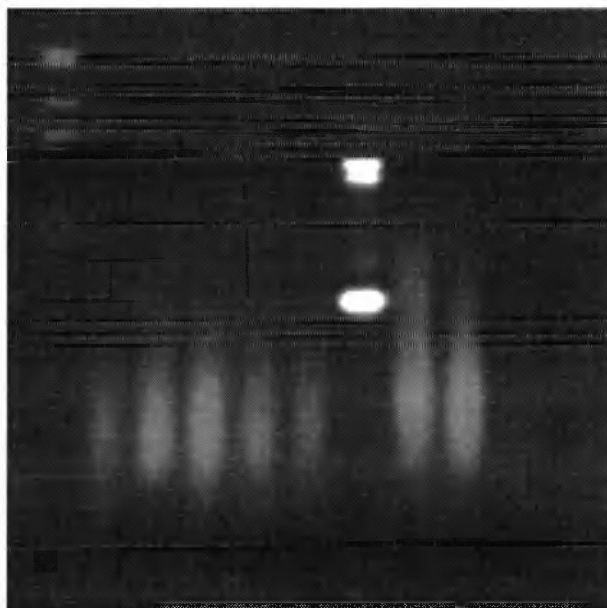
(b) Approximately 1 ng of DNA from chromatin immunoprecipitations (lanes 1 - 5), 1 ng Bluescript (lane 6, RsaI/DraI digested), 1 ng *Drosophila* genomic DNA (lane 7, RsaI digested), or 1 ng sonicated *Drosophila* genomic DNA (lane 8) was ligated to a blunt linker (0.1 μ M) and PCR amplified. 1/20th of the reaction product was then separated on a 1.2 % agarose gel. Lane M is a molecular weight marker, consisting of λ DNA digested with EcoRI and HindIII.

a

	pBS RsaI/DraI			Dm DNA RsaI			Sonicated Dm DNA		
M	1	2	3	4	5	6	7	8	9

**b**

	IPed DNA					Controls		
M	1	2	3	4	5	6	7	8



proportion of the products on an agarose gel (figure 2.3a). DNA was amplified efficiently in all samples using the lowest concentration of linker (lanes 1, 4 and 7); however with increasing concentrations of linker the ligation/amplification became less efficient. The reasons for this are not clear, but the same effect was observed with restriction enzyme-mediated linker-modified PCR (V. Orlando, personal communication). Nevertheless, it is striking that samples containing sonicated genomic DNA amplify almost as efficiently as those containing restriction digested DNA (compare lanes 4 and 7); therefore ligation to sonicated DNA ends is occurring at reasonable efficiency.

1 ng DNA from immunoprecipitation experiments was then ligated under the new conditions and amplified. Amplification, as with the control samples, appeared to be efficient (figure 2.3b). It is however possible that only a very small proportion of immunoprecipitated DNA fragments (e.g. 1 pg) are capable of ligating to linker. Because the PCR amplification is carried out to exhaustion (35 cycles) even such a small DNA population may be capable of amplifying to give the same total amount of DNA as if the starting DNA was 1 ng. In such a scenario the complexity of the starting material able to ligate may be low, and the final amplified probe would therefore not accurately represent the immunoprecipitated material.

To test the complexity of the probes generated by blunt-end linker-modified PCR, DNA from a control immunoprecipitation without antibody was amplified and hybridised to a Southern of the bithorax complex genomic walk (figure 2.4b). Genomic DNA appears to be uniformly represented in the probe, as every restriction fragment in the walk hybridises approximately equally. Restriction fragments hybridising very strongly (in λ clones 4 and 5, and P1 clone 2) correspond to highly repetitive elements. Other fragments, particularly in the P1 clones, apparently give higher than average hybridisation; however these result either from doublets on the agarose gel, or represent elements which are slightly repetitive. In each case, *Drosophila* genomic DNA hybridises in a similar manner (compare figures 2.4b and c). As the λ clones covering *Abd-B* contain less insert DNA than P1 clones few restriction fragments run as doublets upon agarose gel electrophoresis. Therefore the uniform hybridisation of control immunoprecipitated DNA probes is most clearly seen in this region.

2.5. Sonication efficiency of DNA from active and repressed genes

The hybridisation signals produced by a number of independent control immunoprecipitations were quantitated. Plotting the resulting values on a map of the bithorax complex walk confirmed the linearity of the PCR amplification (figure 2.5). Some

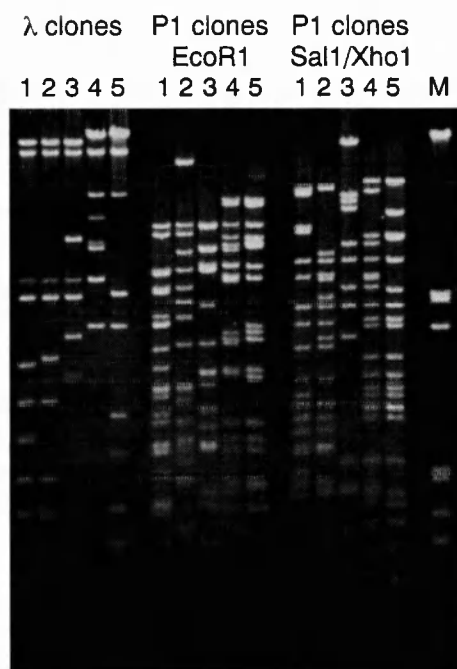
Figure 2.4. Southern hybridisation of control immunoprecipitated DNA to the bithorax complex genomic walk

(a) 1 µg DNA from λ bacteriophage, or 2 µg DNA from P1 bacteriophage was digested with restriction enzymes and separated on a 0.6 % agarose gel. The λ clones cover the distal region of the BX-C; lane 1 (λ8106), lane 2 (λ8099), lane 3 (λ8095), lane 4 (λ8088), lane 5 (λ8083). Lanes 1, 2 and 3 were digested with EcoRI and HindIII, and lanes 4 and 5 with EcoRI. P1 clones cover the remaining 290 kb of the BX-C, and each clone was subjected to two separate digestions, with EcoRI or SalI/XhoI. Distal clones are to the left of the gel, proximal clones to the right; lane 1 (DS04698), lane 2 (DS00846), lane 3 (DS03408), lane 4 (DS03126), lane 5 (DS05563). Lane M is a molecular weight marker, consisting of λ DNA digested with EcoRI/HindIII; the approximate positions of 5.0 kb, 4.2 kb, 2.0 kb and 1.0 kb are indicated adjacent to the marker.

(b) DNA from a control immunoprecipitation was amplified by blunt end linker-modified PCR, and hybridised to a Southern filter of the gel shown in (a). Note that the hybridisation to most restriction fragments is approximately equal, with signal intensity dependent on fragment size. Restriction fragments migrating as doublets show a higher signal intensity, as do those containing repetitive elements; the major repetitive elements are in λ8088/λ8083 (λ clone lanes 4 and 5) and in DS00846 (P1 clone lane 2).

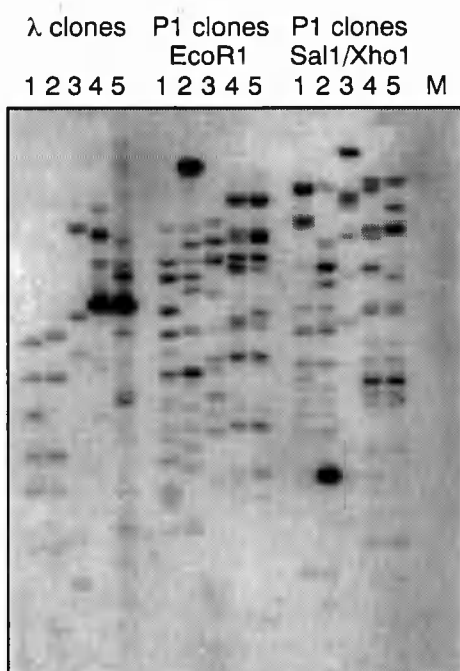
(c) An identical filter hybridised to *Drosophila* genomic DNA. Note that the hybridisation pattern resembles that in (b).

a



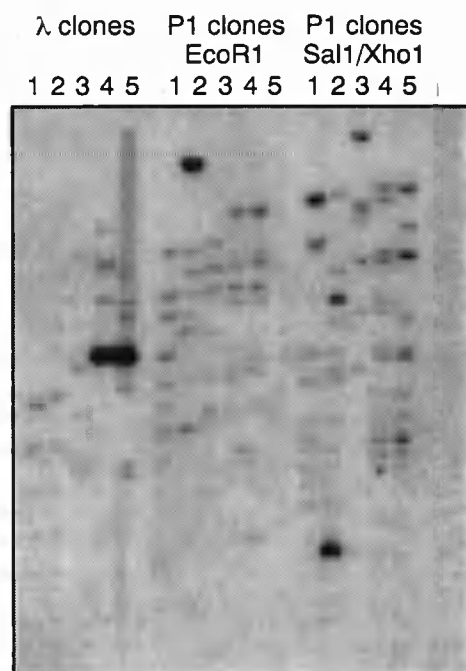
b

Control IP



c

Genomic DNA



fragments are consistently above or below average: this is likely to be caused by sequence-specific differences in amplification inherent to these fragments. However, a mean hybridisation signal could be estimated (solid black line); variation was no more than 30 % above or below this average, except for the repetitive sequences (white bars marked with grey spots below), and this degree of variation between fragment amplification must also be assumed to be present in subsequent experiments.

Interestingly, the immunoprecipitated probe hybridises more strongly to restriction fragments in the *Abd-B* region than to the remainder of the BX-C. In fact, the mean hybridisation signal was 50 % higher than in the *abd-A* and *Ubx* domains, suggesting that DNA from this region is amplified more efficiently. This finding is particularly intriguing in the light of the fact that *Abd-B* is expressed in Schneider cells but the *abd-A* and *Ubx* genes are inactive, and suggests that the chromatin structure may affect the immunoprecipitation result.

As it is already apparent that smaller DNA fragments are amplified more efficiently than larger fragments, differences in DNA size leading to differences in PCR efficiency are a likely cause of the hybridisation differences observed in the BX-C. It is thought that active genes have a more open chromatin structure than repressed genes. Therefore it is possible that the *Abd-B* gene is sonicated more efficiently than the inactive domains of the BX-C, which are more compact and thus more resistant to mechanical shearing.

This hypothesis was tested, by reversing the cross-links of an aliquot of sheared, purified chromatin, without taking it through an immunoprecipitation experiment. 5 µg of this DNA was run on an agarose gel, resulting in a DNA smear ranging from 20 kb to 200 bp, with an average size of 1 kb (figure 2.6a, lane 1). Identical gel strips were blotted, and the filters hybridised to a number of probes originating from genomic regions that are inactive (lanes 2 and 3) or expressed (lanes 4 and 5) in Schneider cells. These probes hybridise as a smear to the cross-linked chromatin, but the smear with the active gene probes is concentrated in the lower molecular weight range. The intensity of signal along the length of the smear was quantitated and plotted against the molecular weight (figure 2.6b). The two active gene probes hybridise to the chromatin smear with an average size of 0.8 kb (grey arrow), whereas the inactive gene probes hybridise with an average molecular weight of 1.2 kb (black arrow). Therefore this experiment confirms that the difference in the level of hybridisation in the BX-C can be accounted for by a difference in sonication, depending on whether the domain is active or repressed.

Figure 2.5. Profile of hybridisation of control immunoprecipitated DNA to the bithorax complex

The BX-C is depicted with proximal to the left and distal to the right. The various regulatory regions of the BX-C (*abx* to *iab-9*) are indicated at the top of the figure. The exon structure of the three homeotic genes is shown in black, and that of other transcripts/ORFs in grey (AHCY, S-adenosyl homocysteine hydrolase-like; GLU, glucose transporter-like; W, low density lipoprotein (LDL) receptor-like repeats; X, serine protease-like; Y, chaperonin-containing t-complex protein-1 γ subunit-like; Z, no-on transient A-like; data from Martin et al., 1995). At the bottom is shown the extent of the P1 or λ clones used in the analysis, together with the restriction enzyme sites for EcoRI (E), SalI (S) and XhoI (X); more detailed restriction site information is in Appendix A1. The traditional map coordinates are shown in bold type (Bender et al., 1983; Karch et al., 1985), and in normal type are the coordinates based on the complete sequence of the BX-C (0 - 340) (Martin et al., 1995).

The hybridisation signals on the Southern filter in figure 2.4b were quantitated, and the resulting values were normalised to account for molecular weight differences and plotted on the map of the BX-C (grey bars). Repetitive elements cannot be accurately quantitated, and are shown as white bars with a grey spot below. Data was taken from at least two immunoprecipitation experiments, carried out on independent cross-linked chromatin preparations. The scale bar to the left indicates relative strength of hybridisation signal, and is in arbitrary units. The mean hybridisation signal is depicted by the thick black line across the profile. Note that in the *Abd-B* domain, the average hybridisation is greater (dashed black line).

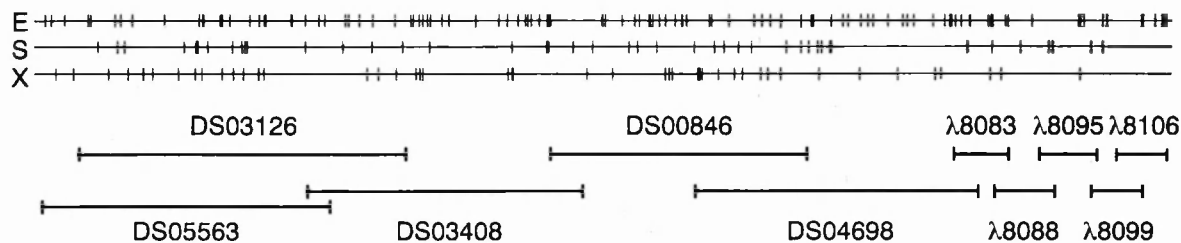
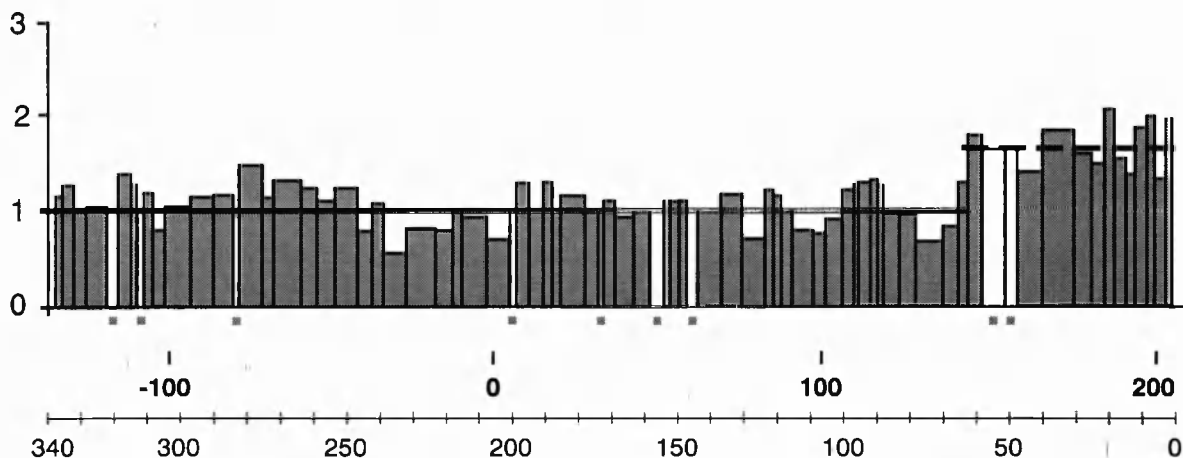


Figure 2.6. The efficiency of sonication of cross-linked DNA depends on gene activity

(a) DNA from an aliquot of sheared, cross-linked chromatin was purified by reversing the cross-links, and 5 µg was separated on a 0.5 % agarose gel (lane 1). Identical gel strips were blotted onto nylon membrane by capillary transfer, and probed with a variety of probes from either active or inactive genes.

Lane 2: probed with a 6.0 kb EcoRI fragment covering the *Mcp* PRE, in the inactive region of the BX-C.

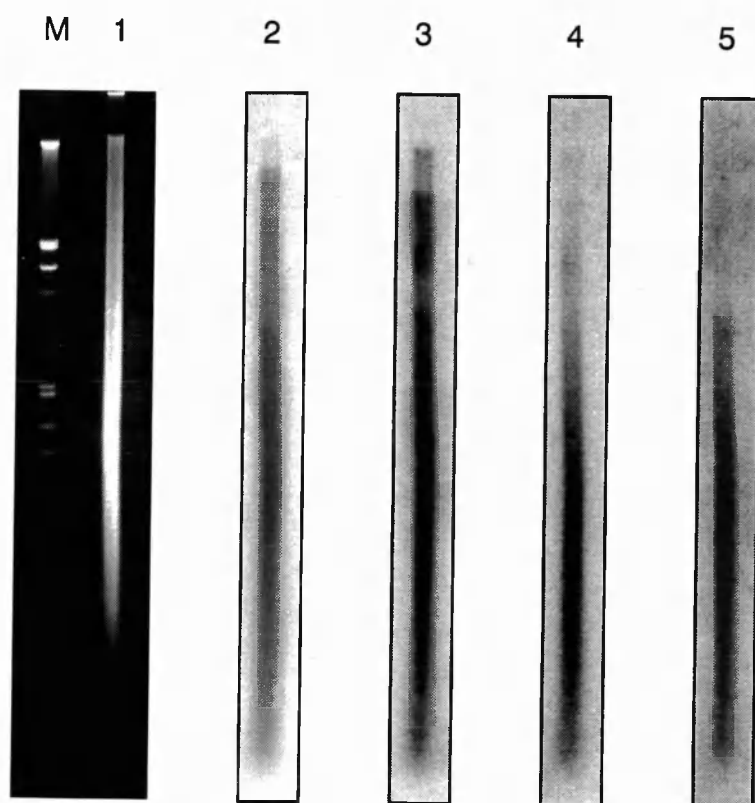
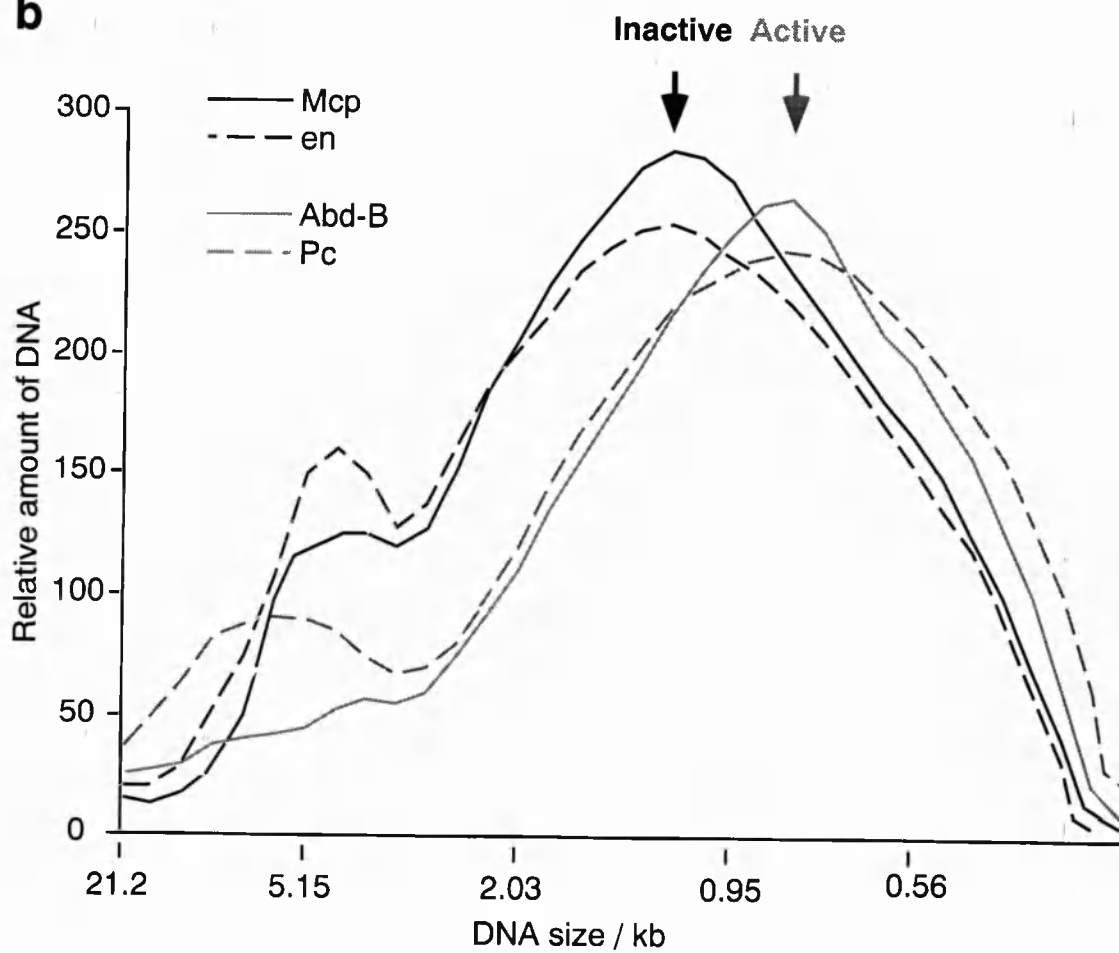
Lane 3: probed with a 4.6 kb EcoRI fragment overlapping the *en* transcription unit, not expressed in Schneider cells.

Lane 4: probed with a 1.5 kb HindIII fragment from the BX-C λ clone 8088, in the expressed *Abd-B* domain.

Lane 5: probed with a 2.4 kb cDNA encoding the *Pc* gene, which is expressed in Schneider cells.

Lane M is a molecular weight marker, λ DNA digested with EcoRI/HindIII.

(b) The intensity of hybridisation of the smears in lanes 2 - 5 was quantitated and plotted with respect to molecular weight. Lines in black show the signal intensity from the inactive gene probes (solid line, *Mcp* probe (lane 2 in panel a); dashed line *en* probe (lane 3 in panel a)), whereas lines in grey show the intensity from expressed gene probes (solid line, *Abd-B* probe (lane 4 in panel a); dashed line *Pc* probe (lane 5 in panel a)). The vertical arrows show the average size for silenced (black) or active (grey) DNA.

a**b**

2.6. Determining the background level

Figures 2.4b and 2.5 show that DNA is isolated non-specifically in control immunoprecipitations, and hybridisation of this DNA to genomic walks can be visualised by long exposures to Phosphorimager screens. Such non-specifically purified DNA is also likely to be present in antibody immunoprecipitations, with DNA specifically associated with the protein of interest enriched several fold. Therefore, not all signals seen by hybridising with antibody immunoprecipitation probes are true signals, a factor which was underestimated in the previous PC binding study (Orlando and Paro, 1993). A detailed slot blot analysis should be carried out to distinguish real enrichments from background. Typically, an equal amount of DNA from plus or minus antibody immunoprecipitations is immobilised on nylon membrane by slot blot, and hybridised to a number of probes derived from the target DNA of interest (for an example, see figure 2.7). The resulting signals are quantitated and the actual enrichment accurately determined. In this manner, the hybridisation of control immunoprecipitated DNA on a Southern of a genomic walk can be translated into a background level in antibody immunoprecipitations: signals seen below this background level of hybridisation are not considered to be enriched.

2.7. Discussion

The formaldehyde cross-linking and immunoprecipitation method has been improved, to allow linear PCR amplification of immunoprecipitated DNA fragments. The validity of this method is further confirmed in chapter 3, by the immunoprecipitation of GAGA factor, a DNA binding protein with a moderately well-characterised DNA consensus binding sequence.

In most experiments carried out using blunt-end linker-modified PCR, the immunoprecipitated DNA amplified uniformly. However on a minority of occasions the amplified DNA had clearly lost complexity, and the probe only hybridised to a few of the expected restriction fragments of a genomic walk (data not shown). This suggests that only a proportion of the immunoprecipitated DNA can be ligated and amplified efficiently, and this amount is close to the acceptable limit in terms of complexity. Therefore, it is important that all immunoprecipitation experiments with a particular antibody are carried out a number of times, to ensure that apparent strong enrichments are reproducible and not caused by fluctuations in complexity. Indeed, the strong enrichments shown in the various immunoprecipitation experiments in later chapters were clearly present in independent experiments. Because of the occasional problems seen with the PCR, it is not clear whether

Figure 2.7. Analysis of plus and minus antibody immunoprecipitated DNA by slot blot

Approximately 100 ng PCR-amplified DNA from both control and PC immunoprecipitations (IPs) was immobilised on nylon membrane by slot blot. Note that panel (a) is a shorter exposure than panels (b) and (c).

(a) The filter was probed with a 4 kb EcoRI fragment from the *abd-A* region of the BX-C (fragment 6 in table 3.1), which is strongly enriched by PC.

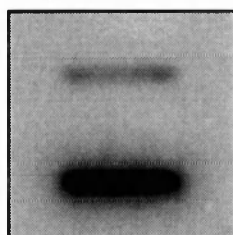
(b) The filter was stripped and reprobed with a 5 kb EcoRI fragment from the *Ubx* region of the BX-C (fragment 7 in table 3.1), which is weakly enriched by PC.

(c) The filter was stripped and reprobed with a 9 kb HindIII-EcoRI fragment from the *Abd-B* region of the BX-C (fragment 2 in table 3.1), which is not enriched by PC.

(d) The filter was finally stripped and reprobed with genomic DNA, as a control for loading on the slot blot. DNA from plus and minus antibody immunoprecipitations hybridise equally to genomic DNA, showing that exactly equal amounts of DNA were fixed on the filter by slot blot.

a

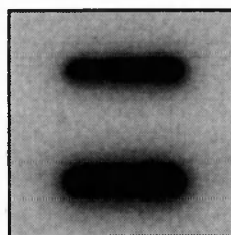
Probe: BX-C
fragment 6



7.5-fold

b

Probe: BX-C
fragment 7



2.0-fold

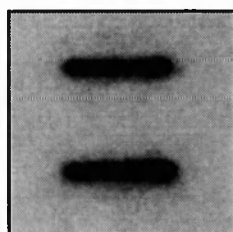
Control IP

PC IP

Enrichment
in PC IP

c

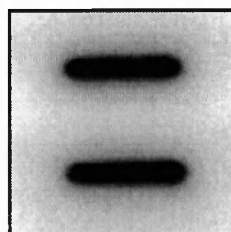
Probe: BX-C
fragment 2



1.0-fold

d

Genomic DNA
probe



1.0-fold
(Loading control)

Control IP

PC IP

Enrichment
in PC IP

this amplification strategy would be appropriate for other organisms with larger genomes, without a corresponding increase in starting material. However, it may be possible to increase the ligation efficiency by using enzymes to repair the sonicated DNA ends.

The background level of hybridisation is greater in active than in repressed genes (figure 2.5). Interestingly, the opposite effect is likely to be observed using the original restriction enzyme-mediated linker-modified PCR strategy. In this procedure, the restriction fragment size affects efficiency of amplification. Furthermore, for efficient exponential amplification any particular DNA fragment must possess two cohesive ends which can be attached to the linker. Therefore, if sonicating cross-linked chromatin results in smaller DNA fragments in active genes, it is less likely that any particular restriction fragment will be preserved with cohesive ends. In this case the amplification of active genes, and the level of hybridisation, will be reduced. Therefore it is possible that the low level of signal observed in the *Abd-B* region by Orlando and Paro (1993) was caused by the methodology.

Using the modified PCR amplification procedure, it has now been possible to map at high resolution the binding sites of PcG and trxG proteins in the BX-C and other target genes (see chapters 3 and 5). Furthermore, this method should be applicable for the analysis of any DNA-binding or chromatin-associated protein over extended genomic regions.

- CHAPTER 3 -

Co-localisation of Polycomb protein and GAGA factor at regulatory elements of the bithorax complex

3.1. Introduction

The data from Orlando and Paro (1993) show that PC is associated with the inactive domain of the BX-C (the *Ubx* and *abd-A* genes), but not with the expressed *Abd-B* gene. However, the results of the previous chapter clearly show that high resolution mapping of PC binding sites was not accurate.

This chapter reports the analysis of PC distribution on the bithorax complex at high resolution, using the improved PCR amplification strategy described in chapter 2. PC is found to be not homogeneously associated with the entire inactive domain, but is highly enriched at discrete sequence elements, which in many cases coincide with characterised PREs. Furthermore, I observe that GAGA factor, a trxG protein, is also constitutively bound to PREs.

3.2. Association of Polycomb protein with the bithorax complex

DNA purified from PC-immunoprecipitated chromatin was amplified by the new and improved blunt end linker-modified PCR strategy and used as a probe against the BX-C genomic walk (figure 3.1a and b). All signals in the Southern hybridisation were then quantitated by Phosphorimager analysis. The resulting values were normalised to account for molecular weight differences between bands and plotted according to their position on the BX-C walk (figure 3.2a).

The level of control DNA hybridisation relative to the hybridisation of PC immunoprecipitated DNA was determined by immobilising 100 ng of control or PC immunoprecipitated probe on a nylon membrane by slot blot, and hybridising with a number of probes from the BX-C (labelled 1 - 9 in figure 3.2a; see table 3.1). Comparison of different fragments allowed the setting of an approximate background level, and below this level fragments are not considered as enriched. As discussed in chapter 2, the background level of hybridisation is higher in *Abd-B* than in the rest of the walk. In addition, repetitive elements hybridise strongly in both PC and control probes, as previously observed.

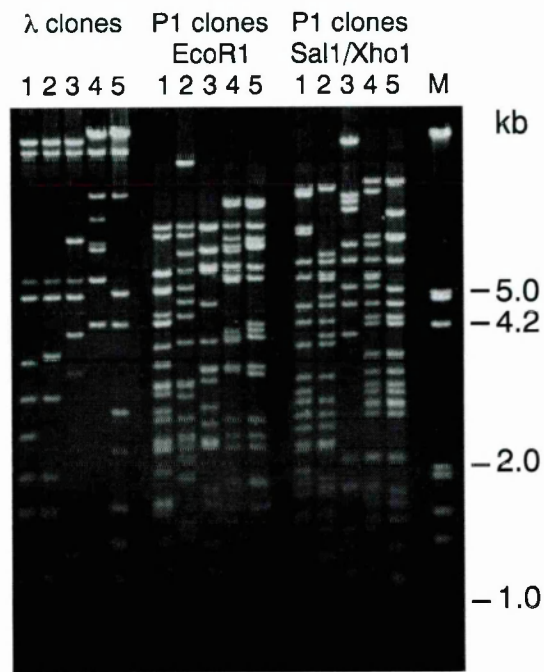
Figure 3.1. Southern hybridisation of Polycomb and GAGA factor immunoprecipitated DNA to the bithorax complex genomic walk

(a) 1 µg DNA from λ bacteriophage, or 2 µg DNA from P1 bacteriophage was digested with restriction enzymes and separated on a 0.6 % agarose gel. The λ clones cover the distal region of the BX-C; lane 1 (λ8106), lane 2 (λ8099), lane 3 (λ8095), lane 4 (λ8088), lane 5 (λ8083). Lanes 1, 2 and 3 were digested with EcoRI and HindIII, and lanes 4 and 5 with EcoRI. P1 clones cover the remaining 290 kb of the BX-C, and each clone was subjected to two separate digestions, with EcoRI or SalI/XhoI. Distal clones are to the left of the gel, proximal clones to the right; lane 1 (DS04698), lane 2 (DS00846), lane 3 (DS03408), lane 4 (DS03126), lane 5 (DS05563). Lane M is a molecular weight marker, consisting of λ DNA digested with EcoRI/HindIII; the approximate positions of 5.0 kb, 4.2 kb, 2.0 kb and 1.0 kb are indicated adjacent to the marker.

(b) Hybridisation of DNA from a PC immunoprecipitation to a Southern filter of the gel shown in (a). Many bands hybridise much more strongly in comparison to hybridisations with control immunoprecipitated DNA (figure 2.4b). Major elements enriched by PC, with reference to figure 3.2a, are a 3.0 kb HindIII fragment (peak A) in λ8106/λ8099 (λ clone lanes 1 and 2), 6.0 kb (peak B) and 7.6 kb (peak C) EcoRI fragments in DS04698 and DS00846 (P1 clone lanes 1 and 2), 1.9 kb (peak D) and 3.9 kb (peak E) EcoRI fragments in DS00846 (P1 clone lane 2), a 3.4 kb EcoRI fragment (peak F) in DS03408 (P1 clone lane 3) and a 2.6 kb EcoRI fragment (peak G) in DS03126 and DS05563 (P1 clone lanes 4 and 5).

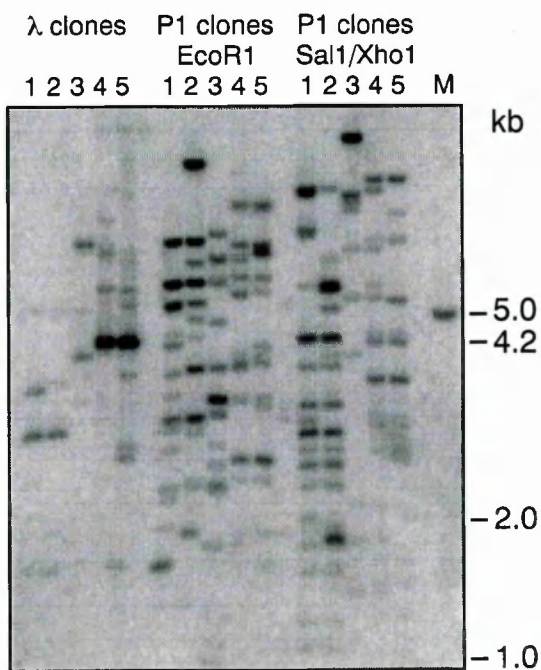
(c) An identical filter to that in (b) hybridised to DNA purified from GAGA factor immunoprecipitations. Note the strong hybridisation of a 4.4 kb EcoRI doublet in DS04698 (P1 clone lane 1) and a 2.6 kb EcoRI fragment in λ8083/DS04698 (λ clone lane 5 and P1 clone lane 1).

a



b

PC IP



c

GAGA IP

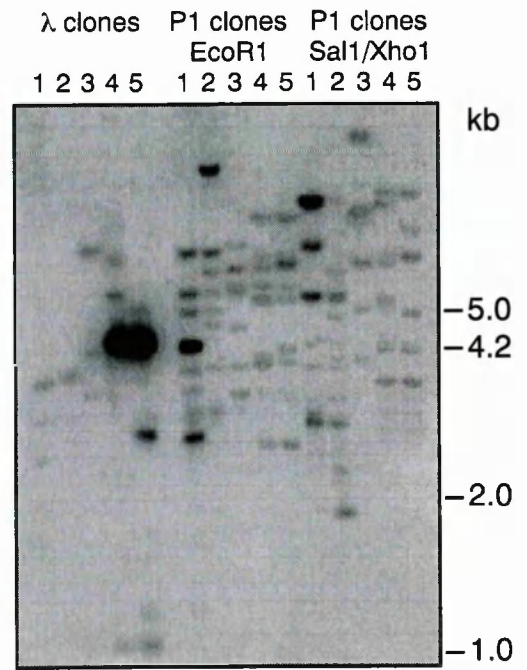


Table 3.1. Slot blot analysis of Polycomb and GAGA factor immunoprecipitated DNA in the bithorax complex

100 ng of control or antibody immunoprecipitated DNA was immobilised on nylon membrane by slot blot, and hybridised to probes 1 - 9 (see also figure 3.2; H, HindIII; E, EcoRI; X, XhoI). The enrichment in antibody immunoprecipitations (IP) is the mean of two experiments, and is calculated with respect to the hybridisation signal in control immunoprecipitations, normalised at 1.0 (n.d., no data). Example slot blots for fragments 2, 6 and 7 are in figure 2.7.

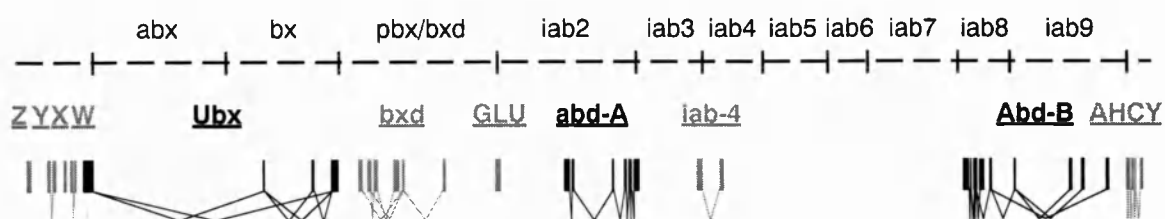
Fragment	Coordinates	Enrichment with respect to control IP	
		PC IP	GAGA IP
1 1956 bp H - E	7680 - 9636	n.d.	0.7
2 9428 bp H - E	31345 - 40773	1.0	1.7
3 3434 bp kb E - E	67073 - 70507	2.0	0.8
4 4344/4389 bp E - E	79681 - 88503	1.4	4.4
5 6741 bp E - E	137414 - 144155	2.3	0.7
6 3870 bp E - E	159944 - 163814	6.5	n.d.
7 4879 bp E - E	227636 - 232519	2.0	0.6
8 5675 bp E - E	284870 - 290545	1.4	0.5
9 4342 bp X - X	307912 - 312254	2.0	0.7

In agreement with the previous analysis, PC is absent from the coding region of *Abdominal-B* (*Abd-B*) (figure 3.2a), which is expressed in Schneider cells (Orlando and Paro, 1993). Significantly, a number of discrete fragments are highly enriched in PC-immunoprecipitated DNA which are not enriched in the mock immunoprecipitation carried out in parallel (compare figures 2.4b and 3.1b). Many of these peak PC binding elements correspond to sequences previously identified as having PRE activity. In particular, these peak PC-binding fragments include the *Mcp* PRE (peak B), the *bxd* PRE (peak F) and the *bx* enhancer PRE (peak G), and overlap with the *iab-3* (peak D) and *iab-2* (peak E) PREs (Busturia and Bienz, 1993; Chan et al., 1994; Qian et al., 1993; Simon et al., 1993). Therefore it is likely that the other peak binding sites for PC also correspond to additional, so far uncharacterised PREs. Interestingly, these results also show that there is one PC peak in each parasegmental regulatory domain. However, PC is not associated with all PRE elements, as poor enrichment is observed at the *Fab-7* PRE (Busturia and Bienz, 1993; Zink and Paro, 1995).

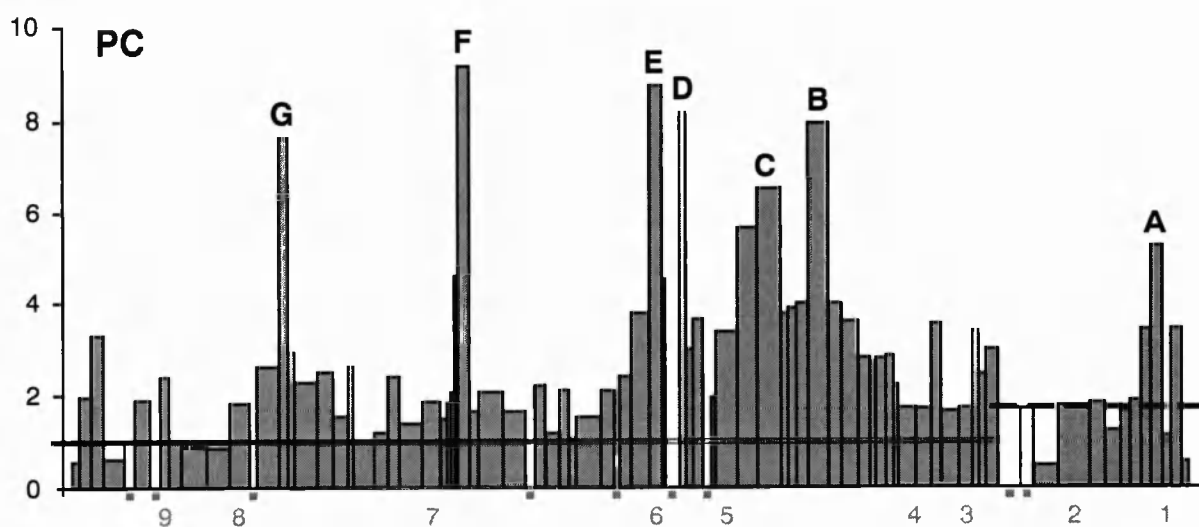
Figure 3.2. Distribution of Polycomb and GAGA factor on the bithorax complex

The BX-C is depicted with proximal to the left and distal to the right. The various regulatory regions of the BX-C (*abx* to *iab-9*) are indicated at the top of the figure. The exon structure of the three homeotic genes is shown in black, and that of other transcripts/ORFs in grey (AHCY, S-adenosyl homocysteine hydrolase-like; GLU, glucose transporter-like; W, low density lipoprotein (LDL) receptor-like repeats; X, serine protease-like; Y, chaperonin-containing t-complex protein-1 γ subunit-like; Z, no-on transient A-like; data from Martin et al., 1995). At the bottom is shown the extent of the P1 or λ clones used in the analysis, together with the restriction enzyme sites for EcoRI (E), SalI (S) and XhoI (X); more detailed restriction site information is in the Appendix A1. The traditional map coordinates are shown in bold type (Bender et al., 1983; Karch et al., 1985), and in normal type are the coordinates based on the complete sequence of the BX-C (0 - 340) (Martin et al., 1995).

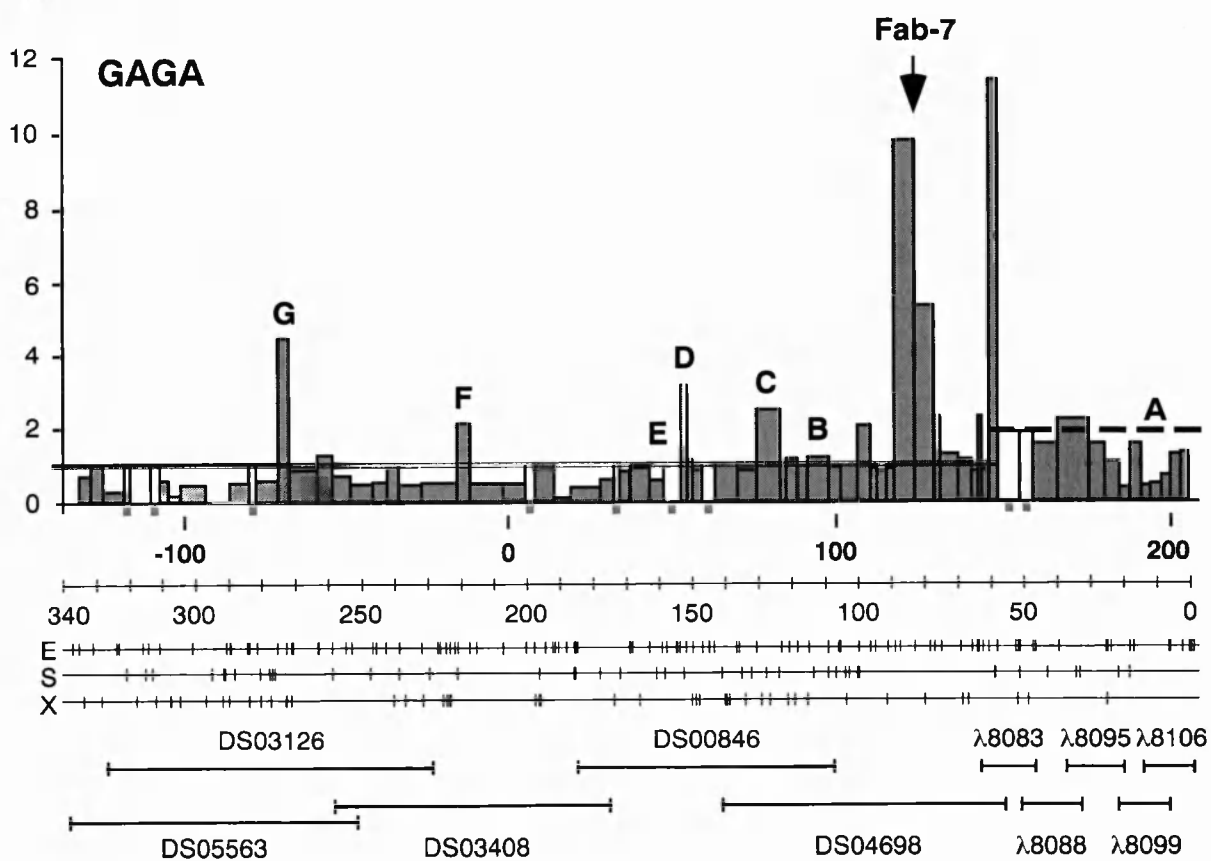
The hybridisation signals on the Southern filters in figure 3.1 were quantitated, and the resulting values were normalised to account for molecular weight differences and plotted on the map of the BX-C. Data was taken from at least two immunoprecipitation experiments, carried out on independent cross-linked chromatin preparations, and results were similar in each experiment. Relative PC (a) and GAGA factor (b) binding is shown as grey bars and repetitive elements are shown as white bars with a grey spot below. Restriction fragments showing strong PC binding are labelled A - G on both profiles, and fragments used for the slot blot analysis to determine background hybridisation (table 3.1) are labelled 1 - 9 in grey, underneath the PC profile. The thick black line across each profile represents the approximate background level of hybridisation, and signals below this line are not considered to be enriched. The scale bar on the left indicates enrichment with respect to this background hybridisation, which is set at 1.0. The background level in the *Abd-B* locus is higher than in the rest of the walk, as discussed in chapter 2 (dashed black line).



a



b



PC does not show a high level association with the entire inactive domain of the BX-C, and in fact many fragments are not enriched. However, DNA adjacent to the peak PC-binding sites is clearly enriched by PC immunoprecipitation. The degree of enrichment appears to decrease with distance from the PRE, indicating that PC associates with a broad domain of 10 - 15 kb, a possible result of spreading from the peak binding site at PREs.

3.3. High resolution mapping of Polycomb immunoprecipitated DNA

Each identified peak of PC binding consists of a single restriction fragment. Therefore an attempt was made to refine to an even higher resolution the sites at which PC is strongly associated. PC binding elements were subcloned into Bluescript (see chapter 7.1 and figure A1 for exact details of the fragments isolated) and digested with appropriate enzymes to yield fragments no greater than 1.2 kb (figure 3.3a). As before, PC or control immunoprecipitated DNA was hybridised to Southern filters of these elements (figure 3.3b and c), and the hybridisation signals quantitated (figure 3.4).

High levels of PC binding in any particular PRE is limited to one or two sub-fragments. In some cases, for example the *Mcp* PRE (figure 3.4b), PC binding is highly restricted, to a single 650 bp DNA fragment. Conversely, binding in the *bx*d PRE (figure 3.4f) is to a larger region of 1.8 kb. Of the known PREs in the BX-C, the *bx*d element has been particularly well characterised (Chan et al., 1994; Chang et al., 1995). Strikingly, in these cross-linking results PC binding coincides with the minimal PRE element.

3.4. Sequence comparison of Polycomb immunoprecipitated fragments

No PcG protein has thus far been demonstrated to bind DNA directly. Therefore, it is not clear if PcG proteins are nucleated at PREs by a specific DNA sequence, or alternatively, if PcG protein complexes recognise a particular DNA conformation at PREs, or interact with other, transient DNA binding proteins. The sequences of the peak PC-binding elements in the BX-C, and of PC binding sites in other loci (see also chapter 5) were compared.

Interestingly, clusters of strong consensus binding sites for GAGA factor, a trxB protein, were found in many PREs (black vertical bars above binding profiles in figure 3.4). In particular, several strong GAGA sites (defined as GA stretches, minimally GAGAG for this analysis) were seen in peak C, peak F (*bx*d PRE) and peak G (*bx* PRE), at the exact site of maximal PC binding. Furthermore, GAGA factor binding sites are seen clustered in peak D in a restriction fragment immediately adjacent to that bound by PC. Peaks A, B (*Mcp*) and E contain no such clusters of GAGA consensus sequences; however GAGA factor has been

Figure 3.3. High resolution mapping of Polycomb and GAGA factor binding at PREs in the bithorax complex

(a) The seven fragments showing peak levels of PC binding (A - G) were subcloned into the Bluescript KS+ plasmid vector, digested with various restriction enzymes and run on a 1.0 % agarose gel. In the cases in which the digests resulted in relatively large fragments, these fragments were purified from agarose gels and further digested. Detailed information about the subclones is given in Appendix A2.

Lane M: Molecular weight marker, consisting of λ DNA digested with EcoRI/HindIII. The approximate positions of 2.0 kb and 1.0 kb are indicated adjacent to the marker.

Peak A: 2982 bp HindIII subclone digested with HindIII/AccI (lane 1); and the two largest fragments further digested; the 1136 bp with AseI (lane 2) and the 912 bp with RsaI (lane 3).

Peak B: 5989 bp EcoRI subclone digested with AccI/PstI (lane 1); the two largest fragments further digested; the 2208 bp with RsaI/XmnI (lane 2) and the 1556 bp with PvuII (lane 3).

Peak C: 7652 bp EcoRI subclone digested with BamHI/PstI/XhoI (lane 1); the two largest fragments further digested; the 1822 bp with HaeII (lane 2) and the 1625 bp with XmnI (lane 3).

Peak D: A 1890/1894 bp EcoRI doublet was resolved by subcloning each into Bluescript. The 1890 bp subclone (lane 1) and the 1894 bp subclone (lane 3) digested with EcoRI; the 1890 bp subclone further digested with XhoI/Eco47III (lane 2) and the 1894 bp subclone digested with PstI (lane 4).

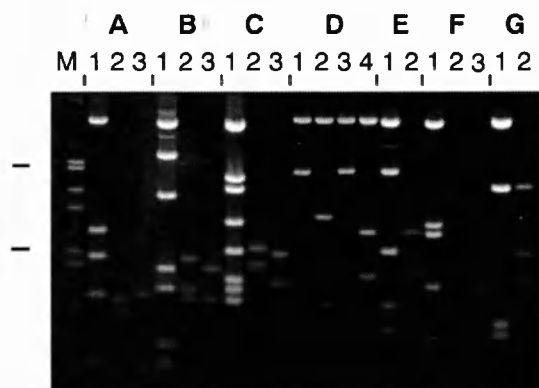
Peak E: A 3870 bp EcoRI subclone digested with HindIII/Sall/ClaI/EcoRI (lane 1); the largest fragment (1883 bp) further digested with HaeII (lane 2).

Peak F: A 3384 bp EcoRI subclone digested with KpnI/PstI (lane 1); the two largest fragments further digested; the 1210 bp with Sau3A (lane 2) and the 1077 bp with NdeI (lane 3).

Peak G: A 2571 bp EcoRI subclone digested with HindIII/EcoRI/Eco47III (lane 1); the largest fragment (1660 bp) further digested with Sau3A (lane 2).

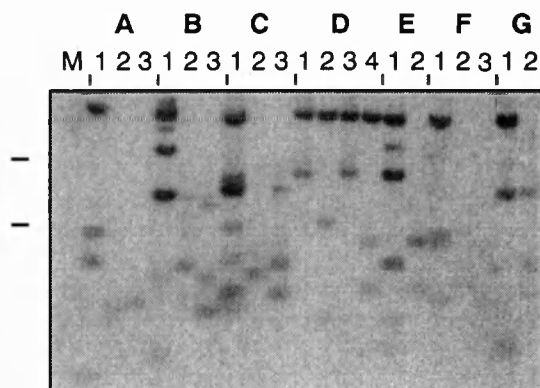
(b) - (d) Southern filters of the gel shown in (a), hybridised to DNA from immunoprecipitations without antibody (b), with PC antibody (c) and with GAGA factor antibody (d).

a



b

Control IP



c

PC IP



d

GAGA IP

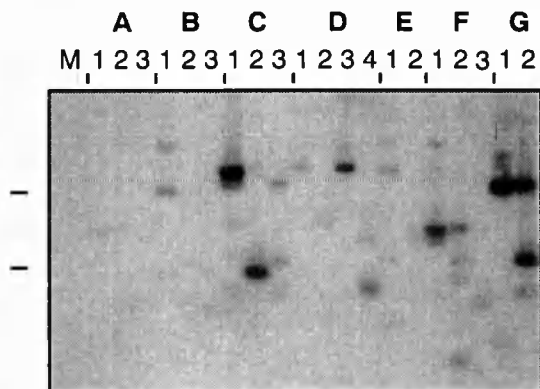
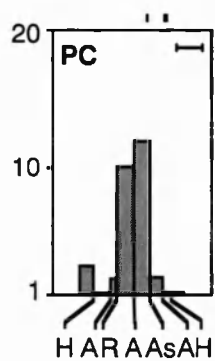
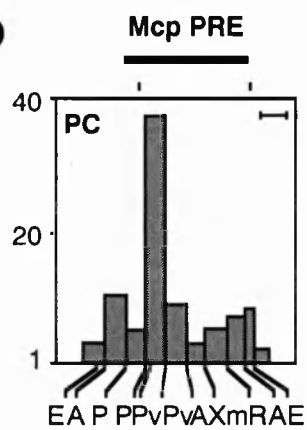
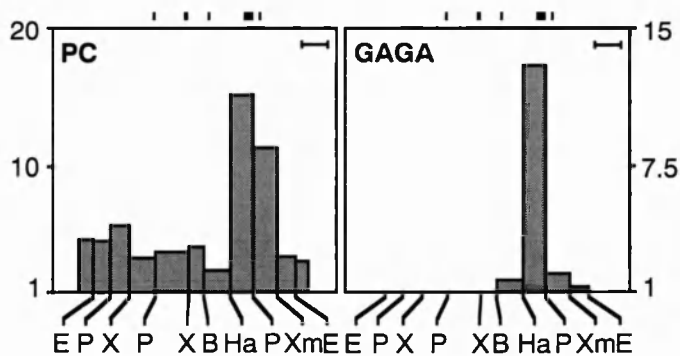
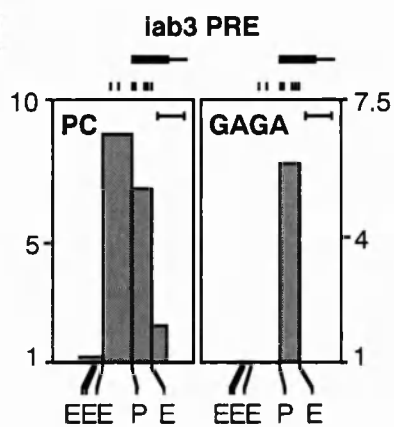
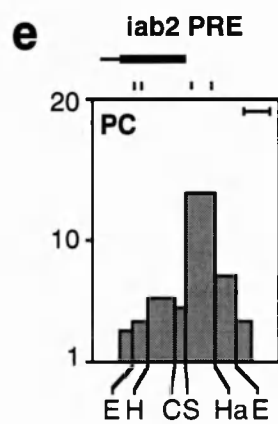
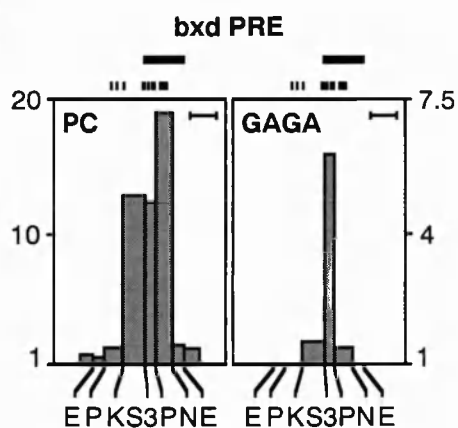
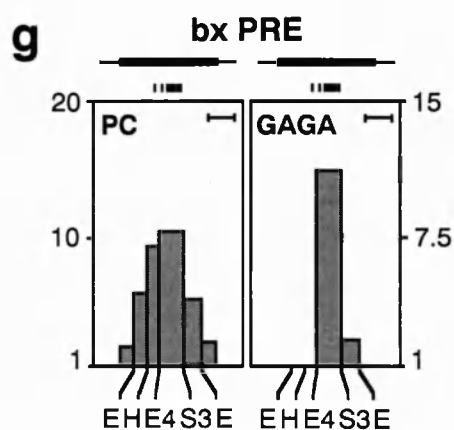


Figure 3.4. Distribution of Polycomb and GAGA factor at PREs

Hybridisation signals from the filters shown in figure 3.3 were quantitated, and normalised to account for molecular weight differences. The total signal for each PRE was used as a guide to normalise the resulting values to the same scale as in figure 3.2. In each case the height of the grey bars indicates relative hybridisation of DNA immunoprecipitated by PC (left) or GAGA factor (right, where applicable) antibodies, and the scale is the same as that in figure 3.2 after subtraction of background hybridisation (set at 1.0). The EcoRI fragments flanking the peak PC binding elements are also shown for comparison. The scale bar in the top right corner of each panel is 1 kb. The vertical black bars above each panel represent consensus GAGA factor binding sites (minimum sequence GAGAG), and the extent of known PREs are shown as horizontal black bars (PREs that are not fully contained within the sequence shown have a black horizontal line extending to the left or right). (a) - (g) represent PC peak binding elements A - G respectively (figure 3.2). Restriction enzymes marked are AccI (A), AseI (As), BamHI (B), ClaI (C), EcoRI (E), Eco47III (E4), HaeII (Ha), HindIII (H), KpnI (K), NdeI (N), PstI (P), PvuII (Pv), RsaI (R), SalI (S), Sau3A (S3), XhoI (X), XmnI (Xm). In each case, proximal is to the left.

a**b****c****d****e****f****g**

demonstrated to bind some promoter elements lacking a strong consensus, so the absence of such a consensus is not necessarily indicative of a lack of GAGA binding (Granok et al., 1995).

3.5. Binding of GAGA factor to consensus sequences in PREs

As GAGA consensus sequences are present at PC binding elements experiments were carried out to investigate if GAGA protein is associated with these elements in Schneider cells. GAGA factor antibodies were obtained which were directed against bacterially-expressed GAGA factor (P. Becker, personal communication). Use of these antibodies in Western blot analysis of nuclear extract from embryos or Schneider cells, or in staining of *Drosophila* polytene chromosomes gave results similar to those observed with previously published GAGA factor antibodies (Raff et al., 1994; Tsukiyama et al., 1994) (G. Cavalli, personal communication).

The association of GAGA factor with a known *in vivo* binding site was analysed first, to test the efficiency of immunoprecipitation with GAGA factor antibodies and the accuracy of mapping GAGA factor binding sites. GAGA factor is constitutively bound to the promoter region of the *hsp26* gene (O'Brien et al., 1995). As expected, chromatin immunoprecipitations specifically enrich restriction fragments from the *hsp26* promoter which contain GAGA consensus binding sites (figure 3.5), but do not enrich for fragments overlapping the transcription unit.

GAGA factor immunoprecipitated DNA was then used as a probe against a Southern of the BX-C walk (figures 3.1c and 3.2b). Two elements hybridise very strongly with the GAGA factor immunoprecipitated DNA probe, relative to the mock immunoprecipitation probe. One is a 2672 bp EcoRI fragment in the *iab-7/iab-8* regulatory region of *Abd-B*, the other a 4344/4389 bp EcoRI doublet which contains the *Fab-7* PRE (Busturia and Bienz, 1993; Zink and Paro, 1995; Hagstrom et al., 1996). Thus the strongest binding of GAGA factor is seen in elements that are not highly enriched for PC. The binding of GAGA factor to the 4344/4389 bp doublet was mapped at high resolution (figure 3.6), and was seen to be limited to a 2.1 kb region, containing clusters of GAGA consensus binding sites. Interestingly, this region contains both the PRE and boundary element function of *Fab-7* (Hagstrom et al., 1996; Zhou et al., 1996; Mihaly et al., 1997).

In addition to the strong binding at *Abd-B* regulatory elements, GAGA factor is associated, albeit at a lower level, with the four PC binding elements which contain GAGA consensus sites (peaks C, D, F and G). High resolution mapping confirmed that in every case

Figure 3.5. Binding of GAGA factor to heat shock genes

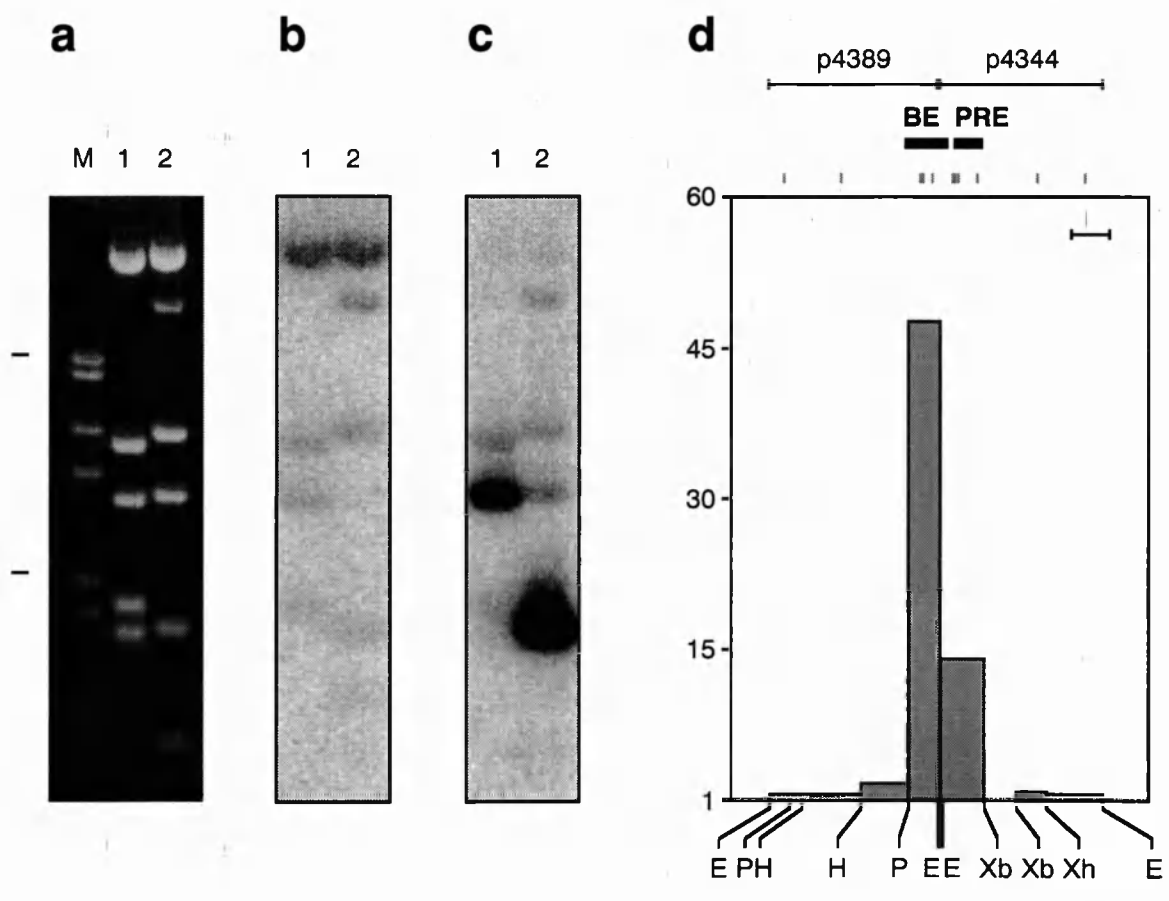
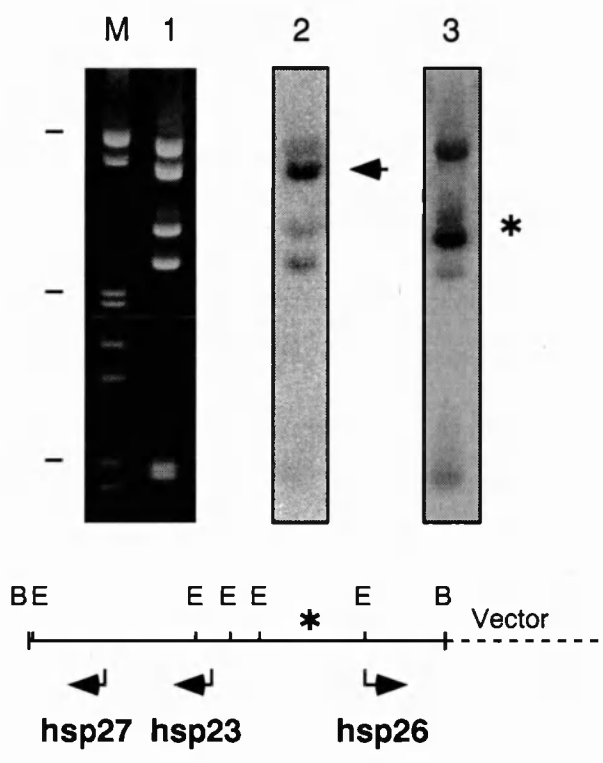
A map of plasmid pJ1 (Craig, 1980), containing *Drosophila* genomic DNA surrounding the *hsp27*, *hsp23* and *hsp26* heat shock genes, is shown with the restriction sites for EcoRI (E) and BamHI (B). Above is an agarose gel containing EcoRI/BamHI digested pJ1 (lane 1), with a marker (M) of λ DNA digested with EcoRI/HindIII. The approximate positions of 5.0 kb, 2.0 kb and 1.0 kb are indicated adjacent to the marker. Lanes 2 and 3 are identical Southern blots of this gel, probed with DNA from control or GAGA factor immunoprecipitations respectively. In the control immunoprecipitation the strongest signal is cross-reaction with the vector (arrow). GAGA factor immunoprecipitations enrich for a 2.8 kb fragment immediately upstream of the *hsp26* transcription start site (the EcoRI site is at +6), indicated by the star. DNA surrounding the *hsp27* gene is also enriched, but not DNA covering the *hsp26* transcription unit or the *hsp23* gene, as previously reported (O'Brien et al., 1995).

Figure 3.6. Binding of GAGA factor to *Fab-7*

(a) The *Fab-7* PRE and boundary element are contained in 2 adjacent EcoRI fragments (4344 and 4389 bp) from clone DS04698. Subclones of these fragments in Bluescript KS+ were digested and separated on a 1 % agarose gel; the 4344 bp subclone digested with XbaI/XhoI (lane 1) and the 4389 bp clone digested with HindIII/PstI (lane 2). Lane M is a molecular weight marker, consisting of λ DNA digested with EcoRI/HindIII; the approximate positions of 2.0 kb and 1.0 kb are indicated adjacent to the marker.

(b) and (c) Identical Southern filters of the gel in (a), probed with DNA from control (b) or GAGA factor (c) immunoprecipitations.

(d) Hybridisation signals from (c) were quantitated and plotted onto an *Fab-7* map (grey bars), after normalising for fragment size and subtraction of the background signal. The extent of the two clones p4344 and p4389 are shown at the top, together with the minimal *Fab-7* PRE and boundary element (BE; Hagstrom et al., 1996). GAGA consensus binding sites (GAGAG) are indicated by grey vertical bars, and the scale bar at the top right of the panel is 1 kb. Restriction enzyme sites are EcoRI (E), HindIII (H), PstI (P), XbaI (Xb), XhoI (Xh).



binding coincides with the presence of clustered GAGA consensus sequences (figures 3.3d and 3.4). The close co-localisation of PC and GAGA factor binding sites indicates that both proteins can be present on the same regulatory elements of silenced genes.

Although PC and GAGA factor show similar binding at PREs, a dramatic difference can be seen between the distribution of the two proteins. Whereas PC protein shows a broad distribution over a few kb surrounding the peak binding site, GAGA immunoprecipitations enrich only those DNA fragments containing GAGA consensus sites, in an all-or-none manner. The difference in the profiles is particularly evident in the high resolution mapping studies (figure 3.4): GAGA factor enriches single fragments to a similar degree as PC, but the neighbouring fragments are not significantly enriched over background. This suggests that PcG complexes form a chromatin structure that is able to spread into DNA adjacent to PREs, whereas GAGA factor does not spread.

3.6. Binding of Polycomb and GAGA factor to the *Antennapedia* P1 promoter

The finding that both PC and GAGA factor are concurrently associated with identical elements is surprising, as these proteins are expected to have opposing functions. As PREs and TREs may be closely related (Chinwalla et al., 1995; Chang et al., 1995), one possibility is that GAGA factor is constitutively bound to elements which are also required for trxG activation. In addition, the presence of GAGA factor at PC-bound PREs may indicate that GAGA factor also has a function in PcG silencing.

The function of PC and GAGA factor was examined at an isolated PRE. A 4 kb element of the *Antp* P1 promoter (from -1.9 to +2.1 kb relative to the transcription start site) was previously shown to possess PRE activity (Zink et al., 1991). GAGA consensus sequences are seen clustered in two distinct regions. One is a region immediately upstream of the transcription start site, consistent with a role for GAGA factor in transcriptional activation at promoters, and the second (containing several strong consensus sequences) is 1.5 - 1.9 kb upstream from the start of transcription.

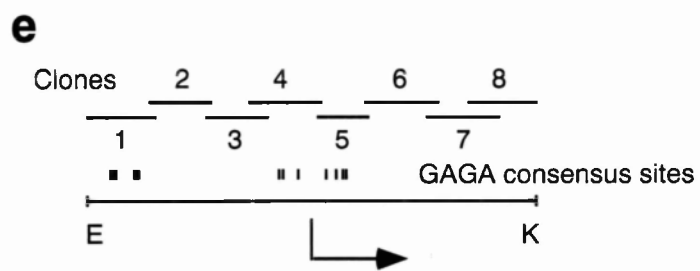
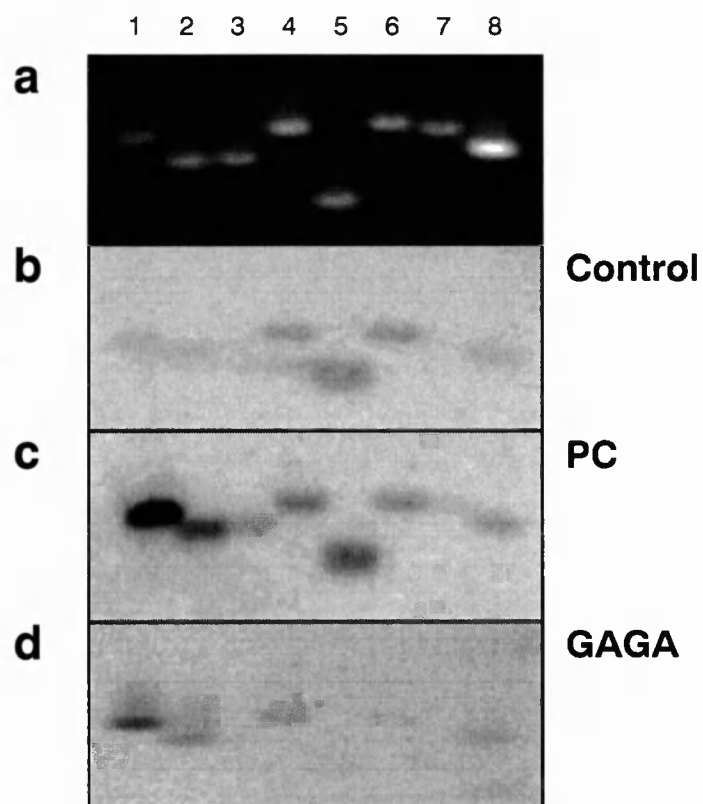
DNA immunoprecipitated with PC antibodies hybridises strongly to the upstream element containing GAGA consensus sequences (figure 3.7), and more weakly to the promoter element. Strikingly, GAGA factor also binds to the upstream element, but does not bind to the consensus sites near the promoter. This result contrasts with what is observed at the heat shock loci, in which GAGA factor binds constitutively to the promoter region.

Figure 3.7. Distribution of Polycomb and GAGA factor at the *Antennapedia* P1 promoter PRE

(a) Eight overlapping subclones of the *Antennapedia* promoter region were digested with restriction enzymes to isolate insert DNA from vector, as described in the Materials and Methods (chapter 7.1), and separated on a 1.2 % agarose gel. Only the insert is shown.

(b) - (d) Southern filters of the gel in (a) hybridised with DNA from control (b), PC (b) or GAGA factor immunoprecipitations (d). Both GAGA factor and PC are enriched in clone 1, and PC is also weakly enriched in the promoter region. The lack of strong binding of PC to fragment 1 in the report by Orlando and Paro (1993) was probably due to NdeII site distribution.

(e) Scheme of the genomic region. A 4 kb EcoRI - KpnI fragment (indicated by E and K respectively) surrounding the P1 promoter was analysed. The start of transcription is shown by the arrow, and the positions of the eight subclones in Bluescript is marked. The positions of the GAGA consensus binding sites are shown as vertical black bars.



3.7. Discussion

GAGA factor: a test of the cross-linking and immunoprecipitation method

The reliability of the new method in the determination of protein binding sites in DNA *in vivo* has been confirmed by immunoprecipitation experiments against GAGA factor. Whilst PC protein has not been demonstrated to bind directly to DNA, GAGA factor is well-characterised in terms of DNA binding, both *in vitro* and *in vivo*. A consensus binding site for GAGA factor, consisting of GAGA repeats, has been proposed. However, GAGA protein can bind to promoter elements lacking such a sequence, and thus any consensus must have additional features (Granok et al., 1995). In fact, GAGA factor immunoprecipitations specifically enrich for DNA containing consensus GAGA binding sites in the BX-C, *Antp* P1 promoter and the *hsp26* promoter. Furthermore, some GAGA sites, such as the promoter-proximal site of *Antp*, are not enriched, arguing against a general reorganisation of chromatin structure during cross-linking allowing artifactual binding to all possible target sequences. Thus, the results produced during this procedure appear to portray accurately the chromatin structure at the time of cross-linking.

Association of Polycomb with PREs in the bithorax complex

Immunoprecipitations against PC protein strongly and specifically enrich for 7 discrete sequence elements in the BX-C. Of these 7 elements, 5 correspond to known PREs, consistent with the idea of PREs being sites for the nucleation of PcG complexes. In particular, the *Ubx* and *abd-A* regions of the BX-C have been extensively analysed for elements conveying PRE activity and it was suggested that there is just one PRE per parasegmental regulatory domain (Chiang et al., 1995). The results shown here are consistent with this hypothesis in that there is one PC binding peak in each regulatory domain that is inactive in Schneider cells.

The two PREs regulating *Ubx*, in the *bxd* and *bx* domains, both correspond to peak PC binding sites (F and G respectively) in this gene. The *bx* PRE has only been analysed as a relatively large element, which contains the PC binding peak (Qian et al., 1993; Simon et al., 1993). However in reporter constructs a minimal 1.6 kb element has been shown to be sufficient for *bxd* PRE activity (Chan et al., 1994) and my high resolution mapping indicates that strong PC binding (in a 1.8 kb element) coincides exactly with this mapped PRE. Moreover, transient transfection assays with reporter constructs encompassing the *bxd* PRE have identified a 440 bp fragment which mediates PC repression (Chang et al., 1995). This fragment is also the most enriched within the 1.8 kb PC binding peak.

PC binding is also seen in regions of the *iab-2* to *iab-5* parasegmental domains (peaks E, D and C respectively). In *iab-2* and *iab-3*, large fragments (11.0 kb and 11.5 kb respectively) immediately adjacent to the peak PC sites have been shown to act as PREs in reporter gene constructs (Simon et al., 1993). The *iab-4* regulatory region has not been extensively analysed for elements showing PRE activity. The finding of a PC binding peak in this domain (peak C) suggests that this element is the *iab-4* PRE. Finally, PC binds strongly to a discrete 623 bp fragment (peak B) contained within the *Mcp* PRE in *iab-5* (Busturia and Bienz, 1993).

The finding that the PC sites in *iab-2* and *iab-3* do not exactly coincide with the elements identified as PREs in reporter gene constructs is puzzling. However, the elements identified here were not also tested in reporter gene assays, and it is likely that they also have the potential to act as PREs. Indeed, it has been suggested that PREs may have multiple sub-elements contributing to the final activity (Pirrotta and Rastelli, 1994), which could be utilised in different spatial or temporal contexts. Therefore in this case it is likely that elements D and E are the preferred site for nucleation of PcG proteins in Schneider cells.

No PREs have been identified so far in the more distal region of the BX-C; however it is likely that peak A, like the other peak PC binding sites, is an *Abd-B* PRE that regulates the γ promoter in *iab-9*. Alternatively, it is also possible that peak PC binding is associated with genes in the BX-C unrelated to the homeotic genes. Sequencing of the BX-C revealed the presence of two ORFs within 10 kb upstream of *Abd-B*; one is a homologue of the human *S-adenosylhomocysteine hydrolase* (*AHCY*) gene, the other shows amino acid identity with a human α -actinin protein (Martin et al., 1995).

No strong PC binding is seen in the regulatory domains *iab-6*, *iab-7* or *iab-8*. In particular, the *Fab-7* PRE, regulating the *iab-7* domain, is poorly enriched in PC immunoprecipitations. The *iab-5* to *iab-9* regulatory regions are each required for modulating differential levels of expression of *Abd-B* in each parasegment of the embryo (Karch et al., 1985). Therefore, the lack of PC binding peaks may indicate that the *iab-6* to *iab-8* domains are involved in positively regulating *Abd-B* expression in Schneider cells (see also chapter 5).

The strong binding of PC at characterised PREs of inactive domains, but not of expressed domains, suggests that the specific association of PC with a PRE is a hallmark of silencing. This would predict an important role for PC in nucleating PcG complexes on the PREs of target genes. Indeed, it was previously shown that tethering a GAL4-PC fusion protein to DNA via artificial GAL4 binding sites was sufficient to recruit an entire PcG protein complex, and to silence neighbouring genes (Müller, 1995). Conversely, PC is

displaced from target PREs when silencing is relieved due to competition from a strong transcriptional activator (Zink and Paro, 1995).

Polycomb protein spreads locally from PREs

It has been suggested that PcG protein complexes repress target loci by forming complexes of heterochromatin-like structures along the chromosome, thus rendering it inaccessible to transactivating factors (Paro, 1990; Orlando and Paro, 1993). However, the results in this chapter demonstrate that PC does not homogeneously cover the entire regulatory regions of the inactive genes of the BX-C. Instead, PC appears to employ specific interactions with discrete regulatory elements in order to silence target genes. However, the distribution of PC in the vicinity of PREs suggests that localised spreading of PcG protein complexes, as opposed to spreading over entire chromosomal domains, may occur to stabilise repression. This apparent spreading is not an artifact due to three-dimensional cross-linking effects, or to poor resolution of the technique, as such a distribution is not seen with GAGA factor immunoprecipitations. On the contrary, such immunoprecipitations enrich for only those DNA fragments containing GAGA consensus binding sequences, to a resolution of less than 1 kb. The difference in the PC and GAGA factor distributions can also not be attributed to a lower sensitivity of the GAGA factor immunoprecipitations, as GAGA factor binding peaks are comparable in intensity to those of PC.

The spreading of PcG protein complexes is reminiscent of silencing events in the yeast *Saccharomyces cerevisiae*. Telomeres are stably repressed by the SIR2, SIR3 and SIR4 proteins, which are thought to form complexes analogous to PcG silencing complexes (Laurenson and Rine, 1992; Pillus and Grunstein, 1995). It has been shown that the SIR proteins spread several kb from their initial nucleation sites, probably by forming stable interactions with histone proteins (Hecht et al., 1996; Hecht et al., 1995). Similarly, limited spreading of PcG protein complexes over a few kilobases may be required for long-term maintenance of silencing.

Spreading of PC is more prominent at some PREs than others (compare the *Mcp* and *bxd* PREs). Thus, DNA flanking the PRE may influence the final distribution of PC. This is in agreement with experiments using a GAL4-PC fusion protein to direct the formation of PcG complexes at artificial GAL4 binding sites, in the absence of a functional PRE. If the fusion protein is removed, the recruited PcG complex is stable throughout many cell divisions if the GAL4 binding sites are flanked by certain *Ubx* sequences, but is not stable if flanked by sequences not normally regulated by the PcG (Müller, 1995). Thus, it was suggested that

"maintenance elements", distinct from PREs, are able to stabilise PcG silencing, by propagating the spreading of repressive complexes (Paro, 1995);

Alternatively, spreading could be due to the binding of PC to weaker PRE elements in the vicinity of the major PRE. Indeed, it was previously suggested that binding of PC to a major PRE may be stabilised by multiple cooperative interactions with weaker PREs (Pirrotta and Rastelli, 1994). These two possibilities cannot be distinguished from the immunoprecipitation data, as there is not sufficient resolution to determine if PC contacts the DNA continuously, or is just associated with multiple, closely-linked DNA elements.

Although the results with PC seem to argue against long-distance spreading as a means of silencing, it is possible that there is a low level of binding over the entire inactive domain, which is below the limits of detection of the method. However, the higher levels of PC at and around PREs indicates that the major role of PC is in the formation of stable complexes at PREs. Other PcG proteins may be more homogeneously distributed in chromatin. It has been predicted that there are 30-40 PcG proteins (Jürgens, 1985), and it would not be surprising if some members of this group, such as PC, are specifically required for nucleating a silencing complex at PREs, whilst other members have more diverse functions.

Binding of GAGA factor to DNA fragments containing PREs

This study shows that in Schneider cells GAGA factor is bound to a number of PREs in the BX-C and to the *Antp* P1 promoter PRE. The strongest GAGA factor binding is in the *Fab-7* element and a more distal 2.7 kb element of the BX-C. Both of these binding sites are located in regulatory domains of the expressed *Abd-B* gene, and are not bound by PC. In addition, a lower level of GAGA factor binding is seen in PREs in the inactive *Ubx*, *abd-A* and *Antp* genes, at which PC is present. Most characterised GAGA consensus binding sites have been located in the promoter elements of target genes (Soeller et al., 1993; Biggin and Tjian, 1988; Gilmour et al., 1989). However, it would be consistent with the role of GAGA factor as a *trxG* gene to find it also associated with more distant regulatory elements.

What is the function of GAGA factor at PC-bound PREs? One possibility is that GAGA factor, which is thought to mediate access of *trans*-acting factors to DNA, may also be required to allow access of PcG complexes to target genes. To investigate a potential role for GAGA factor in PcG silencing, transgenic flies containing the *Antp* P1 promoter and PRE upstream of a *lacZ* reporter gene (Zink et al., 1991) have been analysed. In most lines carrying this construct (pAPT 1.0-5C or pAPT 1.8-20B) PC protein is recruited to the

insertion site in polytene chromosomes and *lacZ* is not expressed in larval salivary glands. Expression of *lacZ* was examined by β -galactosidase (β -gal) staining of salivary glands in lines heterozygous for the reporter gene construct and two different *Trl* alleles, *Trl*^{13C} and *Trl*⁶² (Farkas et al., 1994). No ectopic β -gal staining was observed in a heterozygous *Trl* background, indicating that a reduction in GAGA factor concentration does not relieve PcG silencing (Strutt et al., 1997). It is however possible that the reduction in GAGA factor activity was insufficient for an effect to be apparent in this assay. In fact, *Trl* mutations were shown to enhance the extra sex combs phenotype of heterozygous *Pc* adult males (Strutt et al., 1997). This suggests that there may be a role for GAGA factor in PcG silencing, although the effect may not be direct.

Interestingly, in one line carrying an *Antp* promoter and PRE construct (line pAPT 1.0-79A), PC is not recruited to the insertion site in polytene chromosomes, and the *lacZ* gene is expressed in salivary glands, presumably because of position effects (Zink et al., 1991). *lacZ* expression in salivary glands of this line is much reduced in a *Trl* heterozygous mutant background compared to a wild type background (Strutt et al., 1997). These results provide support for the idea that GAGA factor is present at PC-bound PREs because these elements also act as TREs. However, as GAGA factor consensus sites are present in the *Antp* promoter region as well as the upstream PRE, it is possible that GAGA factor is also bound to these sites in salivary gland nuclei. In this case a reduction in promoter activity due to loss of GAGA factor at these sites could also contribute to the reduced *lacZ* expression.

An overlap between TREs and PREs was previously suggested by the finding that TRX and PC proteins are both bound to a number of identical sites on polytene chromosomes, and can also bind simultaneously to a reporter construct containing the *bxd* PRE in transgenic flies (Chinwalla et al., 1995; Chang et al., 1995). Similarly GAGA factor is also bound constitutively to the *bxd* PRE in the absence of transcriptional activation. The finding that GAGA factor binds to many PC-bound PREs suggests that the overlap between PREs and TREs may be widespread.

The level of GAGA factor binding at PC-bound PREs is similar to that at the *hsp26* promoter in the absence of heat shock induction. In this case GAGA factor is presumed to induce a "poised" state, to allow rapid gene activation under conditions of stress. The mechanism of *trxG* activation at more distant elements is less well understood, but it is possible that constitutive binding of GAGA factor and TRX is required for TRE function. Such constitutively bound proteins may be part of a molecular switch, in which the role of

GAGA factor is to allow access of other trxG proteins and concomitant loss of PcG proteins when activation occurs.

Finally, the binding of GAGA factor to PREs which are silenced by PC is reminiscent of the finding that GAGA factor is bound to other silent DNA domains, the satellite repeats in centromeric heterochromatin (Raff et al., 1994). Mutations in *Trl* have been shown to exhibit defects in the nuclear cleavage cycle, such as failure in chromosome condensation, abnormal chromosome segregation and chromosome fragmentation (Bhat et al., 1996). Similar mitotic defects were also observed in embryos mutant for the heterochromatin protein HP1 (Kellum and Alberts, 1995). It was suggested that GAGA factor is required to organise chromatin structure, to allow access of factors mediating condensation and decatenation in domains which would otherwise be tightly packaged (Bhat et al., 1996). Given the similarity in the mechanisms of PcG and heterochromatin silencing, it is possible that GAGA factor may have a similar role in organising chromosome structure and segregation at PREs in the BX-C.

Role of GAGA factor at the *Fab-7* boundary element

The strong binding of GAGA factor to the two *Abd-B* regulatory elements may represent a different function for GAGA factor than at PC-bound PREs, where lower amounts of GAGA factor are observed. One of these two strong GAGA factor binding sites corresponds to the *Fab-7* boundary element. Similarly, the other strong binding site for GAGA factor is located between the *iab-7* and *iab-8* domains, and could also be a boundary element. Interestingly, two nuclease hypersensitive sites are associated with the *Fab-7* boundary element, and one with the PRE, although it is not known if they are present in all or just a subset of embryonic tissues (Galloni et al., 1993). As the binding of GAGA factor to the *hsp70* promoter leads directly to the formation of hypersensitive sites (Wallrath et al., 1994), it is tempting to speculate that the *Fab-7* hypersensitive sites are also a result of GAGA factor binding.

The binding of GAGA factor to the *Fab-7* boundary is particularly intriguing in view of the fact that *Trl* is also an enhancer of PEV, and is associated with centromeric heterochromatin (Farkas et al., 1994; Raff et al., 1994). Enhancers of PEV have been suggested to function by preventing the spreading of heterochromatin into euchromatic DNA (Reuter and Spierer, 1992). Similarly, the presence of GAGA factor at the *Fab-7* boundary may indicate that it also has a role in blocking the spreading of PcG complexes from the *Fab-7* PRE into *iab-6*.

- CHAPTER 4 -

Association of Polycomb group proteins in a multimeric complex

4.1. Introduction

A number of experiments have indicated that the PcG proteins may act by forming a multimeric protein complex that binds to regulatory sequences of target genes. Double and triple mutant combinations of the PcG genes show a synergistic enhancement of the phenotypes observed with single mutants (Jürgens, 1985), consistent with their participation in a common regulatory structure. In addition, PC and PH have been shown to be members of a multimeric complex which contains 10 - 15 additional components (Franke et al., 1992). Finally, several PcG proteins bind to similar sites on the polytene chromosomes of *Drosophila* larval salivary glands (Franke et al., 1992; Rastelli et al., 1993; Lonie et al., 1994; Carrington and Jones, 1996).

PREs are thought to nucleate the formation of PcG silencing complexes, as isolated PREs in reporter constructs are able to induce the formation of an additional PC binding site at the insertion site of the transposon in polytene chromosomes (Chan et al., 1994; Simon et al., 1993; Zink et al., 1991; Zink and Paro, 1995). The association of PC with PREs may be mediated by the chromodomain, as PC-lacZ fusion proteins which contain chromodomain mutations lose the ability to bind to polytene chromosomes (Messmer et al., 1992).

Moreover, chimaeric HP1 proteins, in which the HP1 chromodomain is replaced with that of PC, are targeted to PC binding sites (Platero et al., 1996). However, PC has not been shown to bind DNA directly, and the chromodomain may in fact associate with DNA indirectly, through protein-protein interactions with other PcG proteins.

In order to investigate the interaction of several PcG proteins both with each other and with target DNA, antibodies against PH and PSC have been generated. By means of immunoprecipitation experiments with these antibodies, PSC is identified as an additional member of the multimeric complex containing PC and PH. Furthermore, mutations in the chromodomain of PC, in addition to causing loss of DNA binding, disrupt the PC/PH protein complex.

4.2. Purification of Polyhomeotic and Posterior sex combs fusion proteins

Antibodies have previously been generated against both PH and PSC (DeCamillis et al., 1992; Martin and Adler, 1993). However, in order to have sufficient antibody to carry out immunoprecipitation experiments it was necessary to produce new antibodies against both proteins. To do this, PH and PSC were expressed in bacteria as fusion proteins, which could then be used for immunisation.

The structure of the *ph* and *Psc* cDNAs is shown schematically in figure 4.1. As both proteins are very large (approximately 200 kD) it was thought to be unlikely that a full-length protein would be efficiently transcribed and translated. Therefore, smaller domains of each protein were selected for expression (see figure 4.1a and 4.1b). Moreover, as the secondary structure of neither PH nor PSC is known, two different domains of each were selected, to optimise the chances of raising antibodies against epitopes that are accessible in full-length protein (potentially in a large complex) *in vivo*. The PH2 fusion protein contains the major glutamine-rich domain of PH, which is similar to that used for generating previously published antibodies (DeCamillis et al., 1992). The PH5 fusion protein is derived from the N-terminus of PH, avoiding the C-terminal SPM domain which is homologous between *ph* and *Scm*, another PcG gene (Bornemann et al., 1996). The PSC7 fusion protein contains the homology region shared with *Suppressor of zeste 2* (*Su(z)2*) and *Psc*; however, previous antibodies against this domain recognised only PSC protein (Martin and Adler, 1993). Finally, the PSC8 fusion protein is derived from sequences immediately downstream of PSC7.

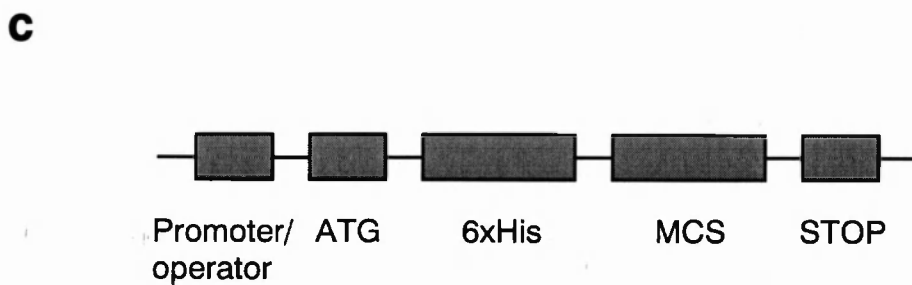
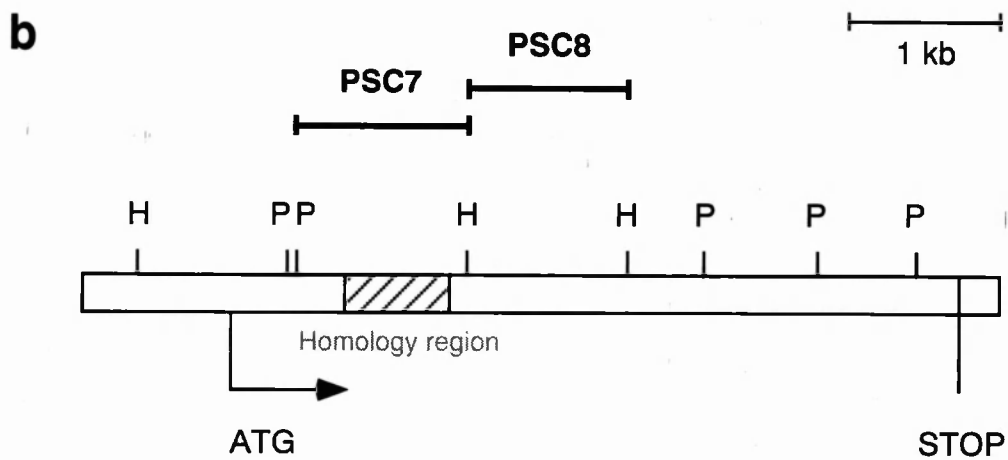
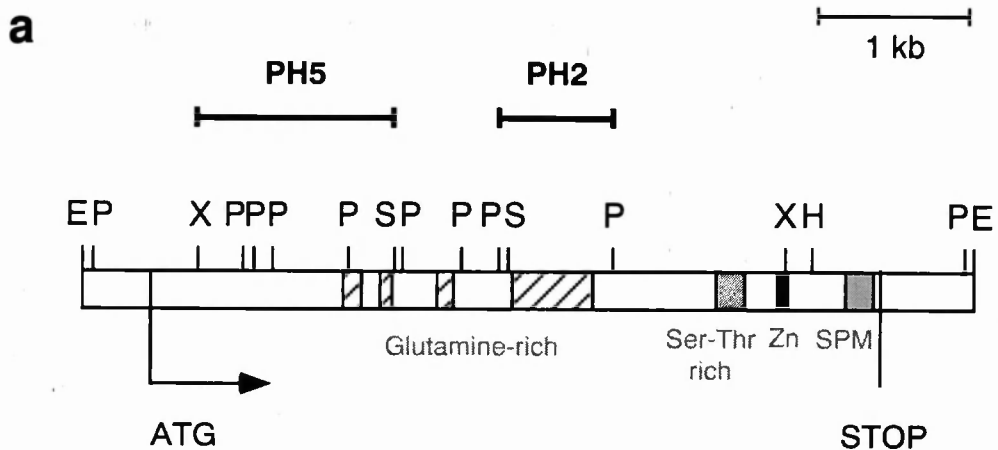
The PH and PSC domains were expressed as fusions with an N-terminal 6 histidine residue tag, by cloning in frame to either pQE (Qiagen) or pRSET (Invitrogen) bacterial vectors (figure 4.1c; see also chapter 7.3). This expression system was chosen for a number of reasons. Firstly, the histidine tag is relatively non-immunogenic, and does not need to be cleaved from the fusion protein prior to immunisation. Secondly, high levels of expression are induced by addition of IPTG to the bacterial culture. Finally, purification of fusion protein is a simple, one step procedure, utilising nickel chelate chromatography. This can be carried out under both non-denaturing and denaturing conditions, depending on whether the fusion protein is soluble or insoluble in bacteria.

PH5 sequences were cloned into a pQE vector. Expression is under the control of the bacteriophage T5 promoter, containing two lac operator sites. Fusion protein constructs were introduced into the *E. coli* strain M15 (Qiagen), which produces high levels of lac repressor from a multi-copy plasmid pREP4. Addition of IPTG to 1 mM inactivates the

Figure 4.1. Cloning of *polyhomeotic* and *Posterior sex combs* cDNAs into expression vectors

(a) and (b) Schematic diagram of the *ph* (a) and *Psc* (b) cDNAs. In each case the translation start site is shown by the arrow (ATG) and the termination site by STOP. The regions PH2, PH5, PSC7 and PSC8 were cloned into pQE (Qiagen) or pRSET (Invitrogen) expression vectors as described in the Materials and Methods (chapter 7.3). In (a) the hatched boxes represent glutamine-rich regions and the dark grey box a serine-threonine rich domain. The black box is a region with zinc finger homology, and the light grey box is the SPM domain, a region of shared homology with *Scm*. Restriction enzyme sites for EcoRI (E), HindIII (H), PstI (P), SalI (S) and XhoI (X) are shown. In (b) the hatched box is the homology region between *Psc* and *Su(z)2*, also containing a RING finger motif. Restriction enzyme sites for HindIII (H) and PstI (P) are shown.

(c) Organisation of the cloning vectors pQE and pRSET. The transcription promoter/operator region is upstream of a translation start site (ATG). Sequences encoding a 6 histidine residue tag (6xHis) are upstream of a multi-cloning site (MCS); three different versions of the pQE and pRSET vectors are available with the MCS in all three reading frames, such that the desired fusion protein can be inserted in frame with the 6xHis tag. Finally a translation stop site is downstream of the MCS.



repressor and allows expression of the fusion protein. Expression of PH5 before and 2 hours after IPTG induction is shown in figure 4.2a. Sufficient fusion protein is produced for it to be visible as an additional Coomassie-stained band in crude bacterial extract (compare lanes 6 and 7).

Fusion proteins PH2, PSC7 and PSC8 were only expressed at very low levels in pQE vectors (data not shown). As an alternative therefore, they were cloned into pRSET expression vectors, in which expression is under the control of the bacteriophage T7 promoter. Constructs were transformed into the *E. coli* strain BL21(DE3) (Studier and Moffatt, 1986), in which the T7 RNA polymerase gene is expressed after IPTG induction. PH2 and PSC8 fusion proteins were strongly expressed 2 hours after addition of 1 mM IPTG (figures 4.2a lanes 2 and 3, and 4.2b lanes 6 and 7); PSC7 was expressed at lower levels (figure 4.2b lanes 2 and 3), but at significantly higher levels than in pQE vectors (data not shown).

All fusion proteins were found to be largely insoluble in bacteria (data not shown) and therefore purification was carried out by nickel chelate chromatography under denaturing conditions (chapter 7.3). Purification was very specific, with negligible contamination of the final product with bacterial proteins; furthermore the purified fusion protein was largely intact, and few degradation products were seen (lanes 5 and 9).

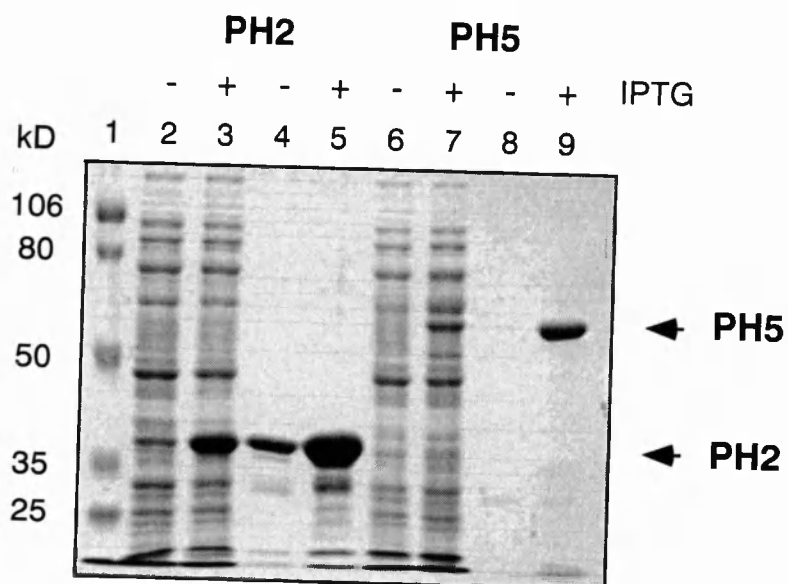
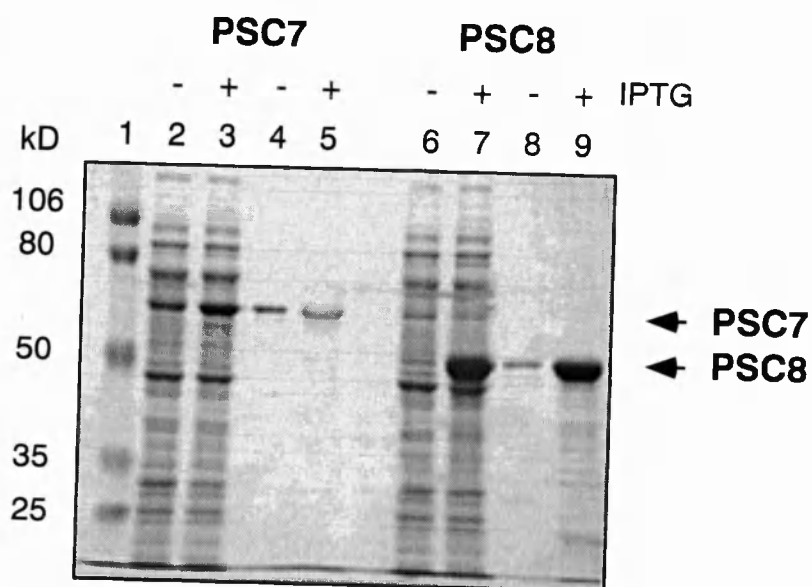
Expression of the PH5 fusion protein in pQE vectors was effectively inhibited in the absence of IPTG (see figure 4.2a lane 8); however expression of the other fusion proteins in pRSET vectors was leaky, with low-level expression even in the absence of IPTG (figure 4.2a lane 4, and 4.2b lanes 4 and 8). This could be a potential problem if fusion protein is toxic to the cells. Indeed the prolonged presence of these fusion protein constructs in BL21(DE3) caused a loss of viability (data not shown), and it was necessary to keep the growth period before induction as short as possible. An alternative strategy would be to transform fusion protein constructs into a strain which does not contain T7 RNA polymerase, and induce expression by addition of a M13 phage expressing the polymerase (Invitrogen).

4.3. Production of Polyhomeotic and Posterior sex combs antibodies

Antibodies against all four fusion proteins were raised in rabbits. After each immunisation serum was tested by Western analysis for a reaction to either fusion protein or full-length protein in nuclear extracts. In all cases a strong reaction against bacterial fusion protein was seen after 1 - 2 boosts, except for PSC7 which required more boosts to cause an immunogenic response. Correspondingly, a strong reaction was detected against *Drosophila*

Figure 4.2. Expression of Polyhomeotic and Posterior sex combs fusion proteins

Coomassie stained 10 % SDS polyacrylamide gels showing the expression of PH and PSC fusion proteins; (a) lanes 2 - 5, PH2; lanes 6 - 9, PH5; (b) lanes 2 - 5, PSC7; lanes 6 - 9, PSC8. Lane 1 is a molecular weight marker (Biorad) with sizes 106, 80, 50, 35 and 25 kD. Lanes 2 and 6 show lysate from 100 µl bacterial culture (solubilised in 8 M urea) without IPTG induction, whereas lanes 3 and 7 show similar cultures which were induced with IPTG. Even in crude lysate the induced fusion protein is visible (arrows). Lanes 4 and 8 show the product of 1 ml uninduced culture, after passage over a nickel chelate column, and lanes 5 and 9 show the fusion protein from 1 ml induced culture after nickel chelate chromatography. PH2, PSC7 and PSC8 fusion proteins (in pRSET vectors) are expressed at a low level in the absence of IPTG induction, whereas expression of PH5 (in a pQE vector) is strictly dependent on the presence of IPTG.

a**PH fusion proteins****b****PSC fusion proteins**

proteins of the expected molecular weight in nuclear extract after 4 - 5 boosts with all fusion proteins except PSC7, which required 7 boosts.

When a strong reaction against full-length protein in *Drosophila* nuclear extract was detected, the entire serum was collected, and affinity purified on a column to which fusion protein was covalently coupled (chapter 7.4). The affinity purified antibodies, the unpurified serum and the preimmune serum were compared by Western blot analysis. In each case the fusion protein (figure 4.3a and b) or full-length protein (figure 4.3c and d) was recognised by serum from immunised rabbits, but not by preimmune serum, and this reaction was maintained after the affinity purification. Both PH antibodies recognise a protein of the same molecular weight in nuclear extract as that recognised by a previously reported antibody (DeCamillis et al., 1992), and the reaction is very specific even with unpurified serum (figure 4.3c). Antibodies against PH5 fusion protein (PH22834 antibodies) also recognise a lower molecular weight protein in nuclear extract (see lanes 6 and 7), which is probably a degradation product of full-length PH, as the signal intensity varied between extracts. Similarly, both PSC antibodies recognise proteins of the same molecular weight as previous PSC antibodies, but they do not recognise a protein corresponding to the homologous *Su(z)2* gene product (figure 4.3d; the predicted molecular weight of *Su(z)2* is 145 kD, smaller than PSC although in fact both PSC and *Su(z)2* from nuclear extracts run slower than predicted). Affinity purification removes immunoglobulins in the serum that cross-react with a number of other proteins in the nuclear extract, but each purified antibody recognises at least one other protein band, again likely to be major degradation products.

Table 4.1. Antibodies against Polyhomeotic and Posterior sex combs fusion proteins

Antibody	Fusion protein	Antibody	Fusion protein
PH24398	PH2	PSC22896	PSC7
PH24414	PH2	PSC22960	PSC7
PH22834	PH5	PSC24376	PSC8
PH22862	PH5	PSC24404	PSC8

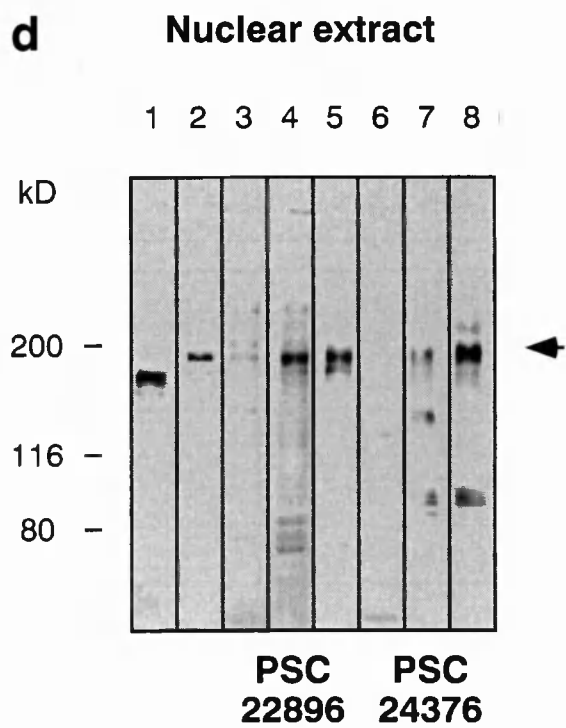
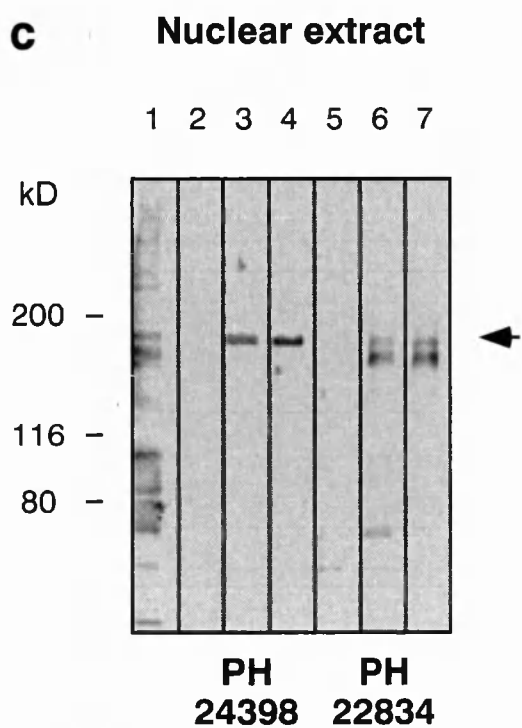
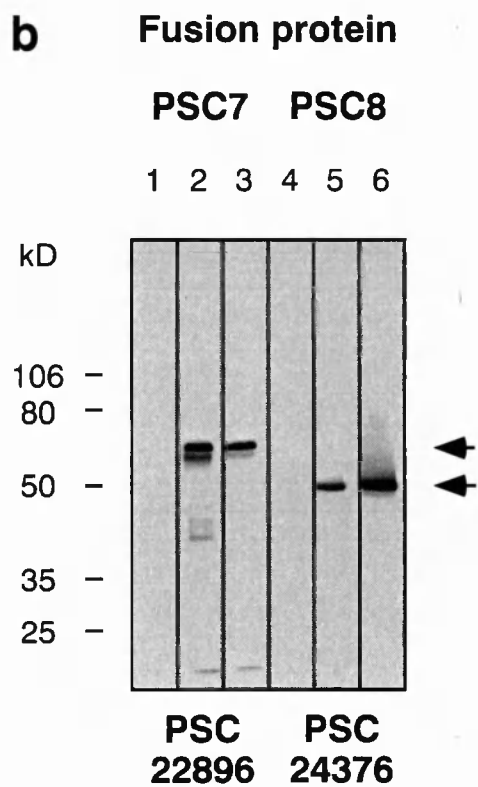
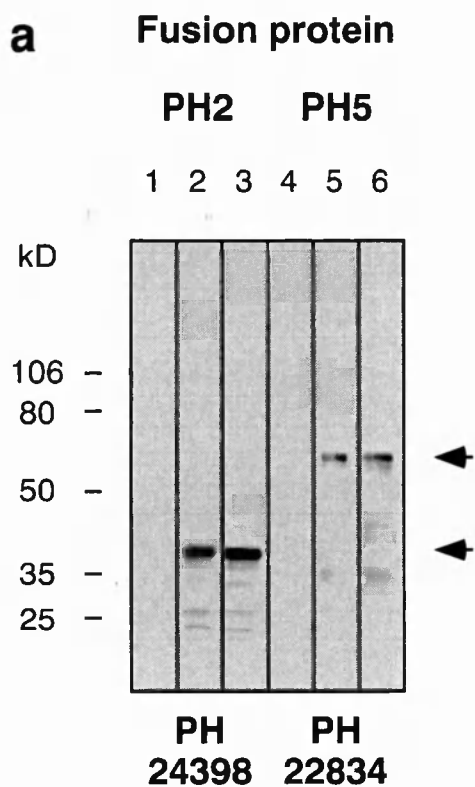
As a final test for specificity, the antibodies were used to stain polytene chromosomes of third instar larvae. Antibodies against both PSC fusion proteins and the PH5 fusion protein bind to polytene chromosomes with a similar distribution to that previously described for PH and PSC (data not shown) (Franke et al., 1992; Martin and Adler, 1993;

Figure 4.3. Western analysis using Polyhomeotic and Posterior sex combs antibodies

(a) and (b) Western blots of 10 % SDS polyacrylamide gels, containing 0.1 µg PH2 (a, lanes 1 - 3), PH5 (a, lanes 4 - 6), PSC7 (b, lanes 1 - 3) or PSC8 (b, lanes 4 - 6) fusion protein, probed with antibodies directed against these domains. Lanes 1 and 4 are probed with the appropriate preimmune serum, lanes 2 and 5 with serum after immunisation of rabbits with fusion protein, and lanes 3 and 6 are probed with the same serum affinity purified; (a) lane 3, PH24398 antibodies; lane 6, PH22834 antibodies; (b) lane 3, PSC22896 antibodies; lane 6, PSC24376 antibodies. In each case, the arrows mark the molecular weight of fusion protein.

(c) and (d) Western blots of 6 % SDS polyacrylamide gels containing 25 µg nuclear extract from *Drosophila* Schneider cells. (c) Lane 1 is probed with antibodies previously generated against PH (DeCamillis et al., 1992). Lanes 2 - 4 are probed with preimmune serum, serum from the same rabbit after immunisation and affinity purified PH24398 antibodies respectively and lanes 5 - 7 with preimmune serum, serum from the same rabbit after immunisation and affinity purified PH22834 antibodies respectively. Both antibodies recognise a protein of the same molecular weight as that recognised by the published antibody (arrow), and PH22834 also recognises a major degradation product in the extracts.

(d) Lane 1 is probed with antibodies against Su(z)2 protein, and lane 2 with antibodies against PSC (a kind gift from M. van Lohuizen). Lanes 3 - 5 are probed with preimmune serum, serum from the same rabbit after immunisation, and affinity purified PSC22896 antibodies respectively, and lanes 6 - 8 with preimmune serum, serum from the same rabbit after immunisation, and affinity purified PSC24376 antibodies respectively. Both antibodies recognise a protein of the same molecular weight as PSC (arrow), but do not recognise Su(z)2. Affinity purification removes many immunoglobulins cross-reacting with nuclear extract proteins, but PSC24376 antibodies still recognise a major protein band at approximately 80 kD.



Rastelli et al., 1993). Antibodies against PH2 fusion protein did not stain polytene chromosomes: it is therefore likely that this epitope is not accessible in fixed tissue. Antibody nomenclature is shown in table 4.1.

4.4. Immunoprecipitation with Polyhomeotic and Posterior sex combs antibodies

Immunoprecipitations were carried out with antibodies directed against all four fusion proteins for a number of reasons. Firstly, it was necessary to test if the antibodies are capable of recognising endogenous protein in its native conformation, as opposed to in a denatured state on a Western blot. Secondly, the buffers used for immunoprecipitation should be tested for compatibility with the antibodies. In particular these antibodies were to be used in immunoprecipitations from cross-linked chromatin (chapter 5) and the stringent buffer conditions used for these experiments were tested in immunoprecipitation experiments from nuclear extract. Finally, these experiments were able to test the hypothesis that PSC is also a member of the multimeric complex containing PC and PH, which was considered likely because of the similarity in the binding sites of all three proteins on polytene chromosomes (Franke et al., 1992; Rastelli et al., 1993). In addition, the mammalian homologues of PC, PH and PSC form a high molecular weight protein complex (Alkema et al., 1997).

Antibodies against all four fusion proteins efficiently immunoprecipitated endogenous protein from nuclear extracts under the buffer conditions used for chromatin immunoprecipitations. Immunoprecipitation experiments with PH24398 antibodies are shown in figure 4.4a. These antibodies specifically purify PH protein when compared to control immunoprecipitations with no antibody (figure 4.4a lanes 1 - 3). Moreover, the protein recognised by the other batch of PH antibodies (PH22834) is also strongly precipitated by PH24398, confirming that they are likely to be recognising the same protein (lanes 4 and 5). In addition, both PC and PSC proteins are co-immunoprecipitated with PH (lanes 6 - 9).

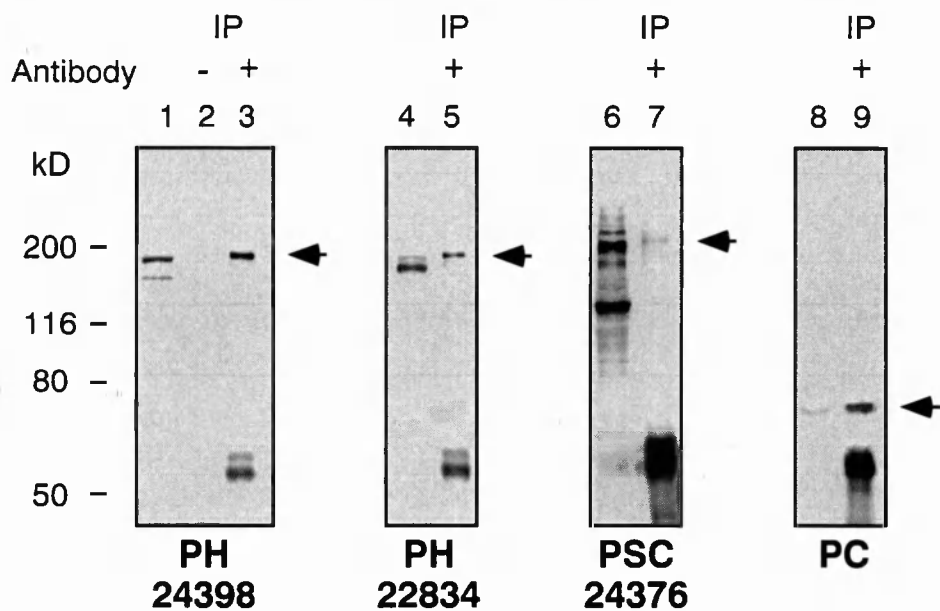
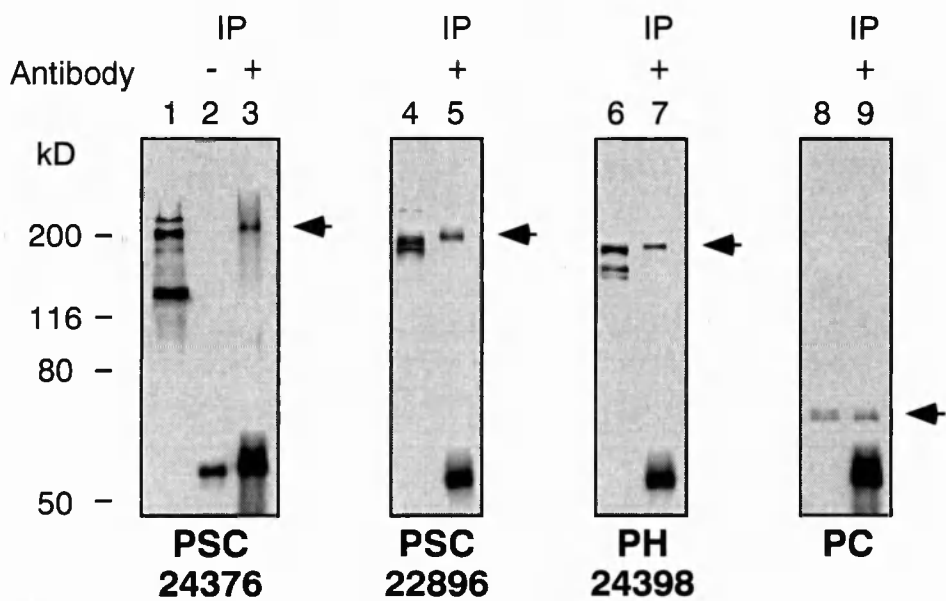
Similarly, both PH and PC are co-immunoprecipitated with PSC antibodies (figure 4.4b). Identical results were observed whichever PH and PSC antibody was used for the immunoprecipitation, and thus these results confirm that PSC is an additional member of the multimeric complex containing PC and PH.

Figure 4.4. Co-immunoprecipitation of Polycomb, Polyhomeotic and Posterior sex combs proteins

25 µg nuclear extract (lanes 1, 4, 6 and 8) or immunoprecipitated material resulting from 75 µg extract (lanes 2, 3, 5, 7 and 9) was separated on 8 % SDS polyacrylamide gels and analysed by Western blotting. Filters were incubated with either PH, PSC or PC antibodies. In each case lane 2 is a control immunoprecipitation without antibodies, and the positions of various molecular weight markers are shown to the left.

(a) Immunoprecipitation (IP) with PH24398 antibodies; Western blots probed with PH24398 (lanes 1 - 3), PH22834 (lanes 4 and 5), PSC24376 (lanes 6 and 7) or PC (lanes 8 and 9) antibodies. Proteins recognised by the two PH antibodies are strongly precipitated by PH24398 (lanes 3 and 5, arrows), confirming that the antibodies recognise the same protein, and PSC and PC are co-precipitated (lanes 7 and 9, arrows).

(b) Immunoprecipitation (IP) with PSC24376 antibodies; Western blots probed with PSC24376 (lanes 1 - 3), PSC22896 (lanes 4 and 5), PH24398 (lanes 6 and 7) or PC (lanes 8 and 9) antibodies. PSC24376 strongly precipitates proteins recognised by both PSC24376 and PSC22896 (lanes 3 and 5, arrows), and PH and PC are co-immunoprecipitated (lanes 7 and 9, arrows).

a**PH24398 immunoprecipitation****b****PSC24376 immunoprecipitation**

4.5. Co-immunoprecipitation of Polycomb-lacZ fusion proteins

The chromodomain of PC and a C-terminal sequence are the only PC domains conserved with the murine homologue M33 and the *Xenopus* homologue XPc, suggesting that these two domains are likely to be important for PC function (Pearce et al., 1992; Reijnen et al., 1995). Indeed, a number of mutations affecting PC function have been identified in both the chromodomain and the C-terminal domain (Messmer et al., 1992; Franke et al., 1995). The chromodomain of PC was previously shown to be required for association of PC with its binding sites on polytene chromosomes (Messmer et al., 1992). The exact molecular role of the C-terminus is not known, although it was found to be necessary for the repression of reporter gene constructs (Bunker and Kingston, 1994; Müller, 1995). The domains of PC which are important for interacting with other PcG proteins was examined by performing co-immunoprecipitations in nuclear extracts from transgenic flies containing PC-lacZ fusion proteins. Three fusion proteins containing C-terminal lacZ fusions were tested (figure 4.5): full-length PC (1 - 390 PC-lacZ; S. Messmer, unpublished data), PC lacking the C-terminal 124 amino acids (1 - 266 PC-lacZ) (Messmer et al., 1992) and finally PC with a chromodomain deletion (1 - 266 PC Δ 42-65-lacZ) (Messmer et al., 1992). All three fusion proteins still contain two blocks of histidine repeats which were shown to exert a minor influence on binding of PC to polytene chromosomes (Franke et al., 1995).

PH antibodies immunoprecipitate PH protein from embryonic nuclear extracts of all three transgenic lines (figure 4.6a). Westerns of immunoprecipitated material were then probed with antibodies against β -galactosidase (figure 4.6c), revealing that fusion protein is efficiently co-immunoprecipitated with PH from both 1 - 390 PC-lacZ and 1 - 266 PC-lacZ nuclear extracts. As fusion protein lacking the C-terminus can form complexes as efficiently as full-length PC, it can be concluded that the C-terminus is not required for the PC-PH interaction. However, a much reduced amount of fusion protein is precipitated from embryonic extracts of transgenic flies in which the chromodomain of the fusion protein is mutated (figure 4.6c, lane 6), even though the immunoprecipitation efficiency is greatest in this line (compare with figure 4.6a, lane 6). This suggests that the chromodomain of PC is required for PC to interact with PH.

As a control, Western filters were also probed with PC antibodies: in every case PC protein is co-immunoprecipitated, as expected (figure 4.6b; in particular note that endogenous PC is co-precipitated in line 1 - 266 PC Δ 42-65-lacZ). PC antibody also recognises PC-lacZ fusion proteins in nuclear extract, but not in immunoprecipitated material as the sensitivity is too low.

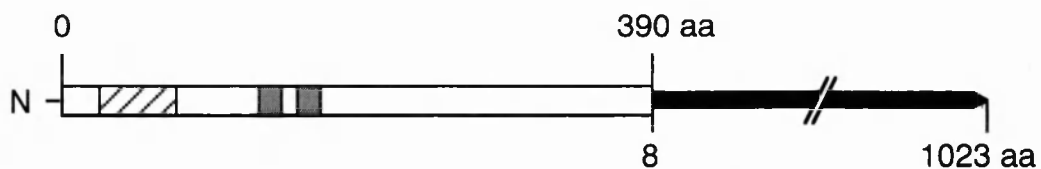
Figure 4.5. Polycomb-lacZ fusion protein constructs

(a) 1 - 390 PC-lacZ. cDNA encoding full-length PC protein (390 amino acids, white) is fused upstream of the lacZ gene (encoding amino acids 8 - 1023, black). The position of the chromodomain is shown by the hatched box (amino acids 26 - 73) and of two histidine repeat stretches by the grey boxes.

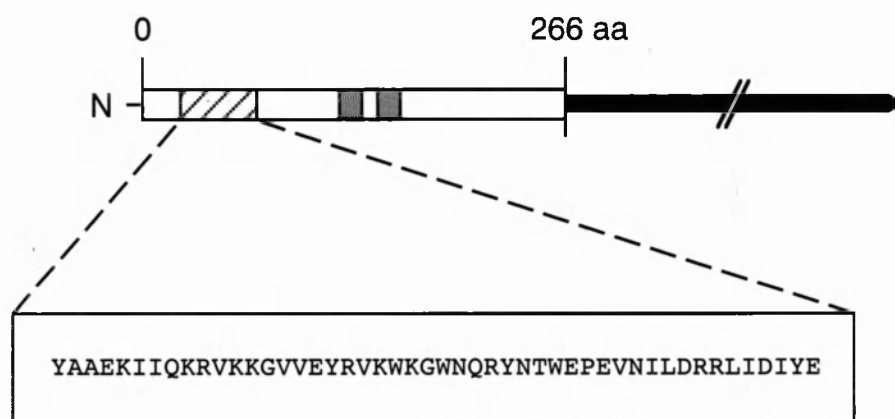
(b) 1 - 266 PC-lacZ. 2/3 PC protein (amino acids 1 - 266) fused upstream of lacZ. The sequence of the wild-type chromodomain is shown below.

(c) 1 - 266 PC Δ 42-65-lacZ. 2/3 PC protein fused upstream of lacZ; the PC protein has a deletion in the chromodomain from amino acids 42 - 65 (shown below).

a 1-390PC



b 1-266PC



c 1-266PC Δ 42-65

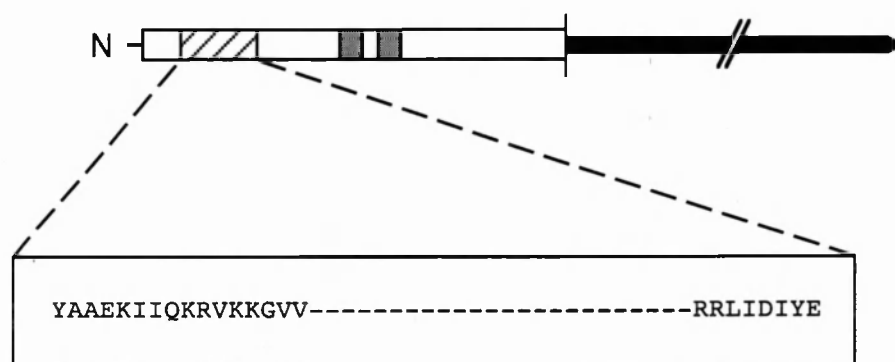


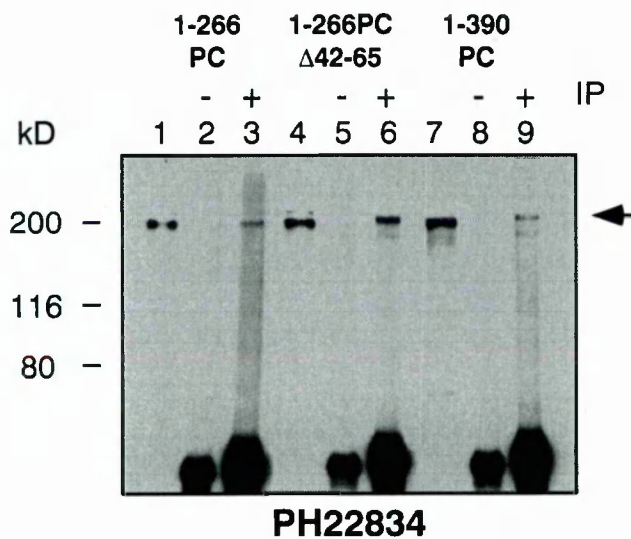
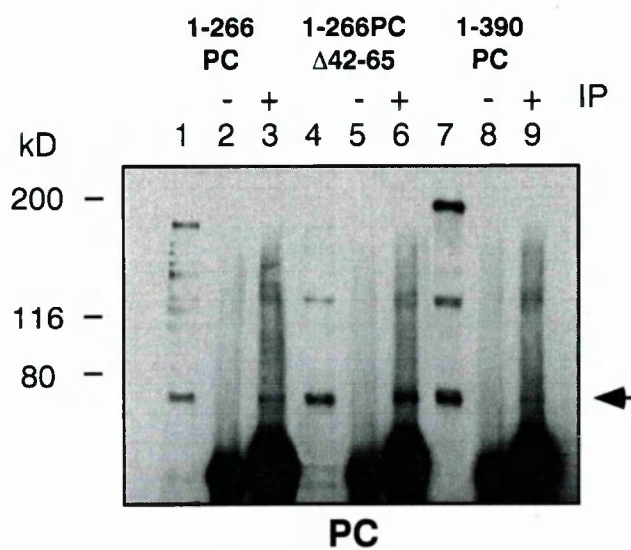
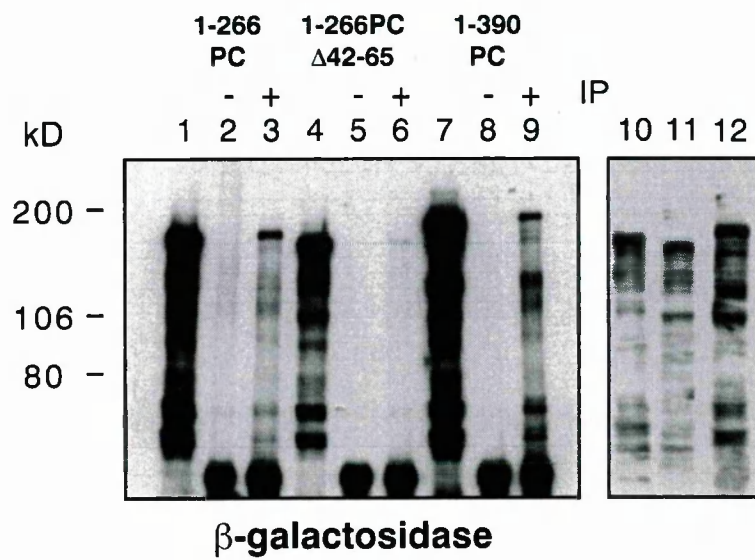
Figure 4.6. Co-immunoprecipitation of Polycomb-lacZ fusion proteins from nuclear extracts of transgenic flies

25 µg nuclear extract (lanes 1, 4 and 7), or 75 µg extract immunoprecipitated with PH antibodies (lanes 3, 6 and 9) or no antibody as a control (lanes 2, 5 and 8) was separated on 8 % SDS polyacrylamide gels and Western blotted. After incubating with PH, PC or β -galactosidase antibodies, the filters were developed by ECL (Amersham). Extract in lanes 1 - 3 is from 1 - 266 PC-lacZ flies, in lanes 4 - 6 from 1 - 266 PC Δ 42-65-lacZ flies, and in lanes 7 - 9 from 1 - 390 PC-lacZ flies. To the left is shown the positions of a molecular weight marker.

(a) Filter probed with PH antibodies; the arrow shows the position of PH protein. Signal is seen in the PH immunoprecipitations (+ IP; lanes 3, 6 and 9) but not in the control immunoprecipitations (-IP; lanes 2, 5 and 8), as expected.

(b) Filters probed with PC antibodies; endogenous PC is at 65 kD (arrow). The higher molecular weight bands are the PC-lacZ fusion proteins and degradation products. PC-lacZ fusion proteins are not recognised by PC antibodies in the immunoprecipitation lanes (3, 6 and 9) because the sensitivity is not high enough (compare intensity of PC-lacZ signal in panel (c)).

(c) Filters probed with β -galactosidase antibodies. Lanes 10, 11 and 12 are a shorter exposure of lanes 1, 4 and 7 respectively; in each case the highest molecular weight band is PC-lacZ fusion protein, which is relatively unstable as many lower molecular weight degradation products are seen. β -gal signals in nuclear extract and PH-immunoprecipitated extract were quantitated by comparison to a dilution series of fusion protein (not shown); 1 - 266 PC Δ 42-65-lacZ fusion protein is immunoprecipitated at only 5 - 10 % the efficiency of the other fusion proteins.

a**b****c**

4.6. Discussion

The Polycomb group protein multimeric complex

The co-immunoprecipitation experiments presented in this chapter demonstrate that PSC is an additional member of a common multimeric complex that contains PC and PH. It is likely that the remaining components are also PcG proteins: for example PCL and E(Z) proteins also bind to similar sites on polytene chromosomes (Carrington and Jones, 1996; Lonie et al., 1994). Fewer E(Z) binding sites were observed than PC sites, but it is possible that this is a problem of sensitivity of the antibody. In fact, temperature sensitive alleles of *E(z)* cause loss of PC binding to most of its sites at the restrictive temperature, suggesting that E(Z) is present but at levels below the limit of detection (Platero et al., 1996).

The product of the *Scm* gene is also a good candidate for being a member of a PcG protein complex. It contains a domain of homology to *ph* (the SPM domain), which was suggested to be a potential dimerisation domain, such that SCM could form homodimers with itself, or heterodimers with PH (Bornemann et al., 1996). Therefore, this molecular data conforms with the hypothesis that the PcG proteins act together as a silencing complex.

The role of the Polycomb chromodomain

My results show that PC fusion proteins lacking the C-terminal domain can efficiently compete with endogenous PC for incorporation into a multimeric protein complex that includes PH. Conversely, fusion proteins with a deletion in the chromodomain are not efficiently incorporated into a PH-containing complex. It has previously been shown that the fusion protein with the chromodomain deletion is also unable to associate with target genes on polytene chromosomes, whereas the C-terminal domain is dispensable for this process (Messmer et al., 1992). Taken together, these results suggest that the primary role of the chromodomain is to mediate PC binding to other members of the PcG, and these other proteins would interact more directly with the chromatin fibre. A role for the chromodomain in protein-protein interactions was previously proposed from recent experiments in which the PC chromodomain present in a chimaeric HP1 protein was shown to be sufficient for recruitment of the chimaeric protein to PC-binding sites in polytene chromosomes (Platero et al., 1995). Interestingly the chimaeric protein retained the ability to interact with its normal heterochromatin target sites, and PSC and endogenous PC were ectopically recruited to these sites (Platero et al., 1996).

The fact that the chromodomain deletion does not completely abolish association with PH suggests that other domains in PC may also contribute to protein-protein interactions. Furthermore, the interaction between the PC chromodomain and PH is not necessarily direct. Instead, it is possible that mutations in the chromodomain cause a general destabilisation of protein-protein interactions within the PcG protein complex. Indeed, a number of experiments have been reported in which mutation of a PcG gene disturbs binding of other PcG proteins to chromosomes. Firstly, mutations in the PC chromodomain disrupt, at least partially, the characteristic distribution of PH in embryonic nuclei (Franke et al., 1995). In addition, disruption of the PcG protein complex with a temperature-sensitive allele of *Enhancer of zeste* (*E(z)*) caused loss of binding of both PC and PSC to most polytene chromosome sites (Rastelli et al., 1993; Platero et al., 1996).

- CHAPTER 5 -

The Polycomb group proteins form complexes of differential composition at different target genes

5.1. Introduction

Although PC, PH and PSC have been demonstrated to be members of a multimeric complex, it is not known how this complex functions *in vivo*. In the classical heterochromatin packaging model for PcG function, the PcG proteins were envisaged as forming a single, discrete complex: reiterated PcG protein units would then spread cooperatively along the chromosome (Locke et al., 1988; Paro, 1990). If this hypothesis is true, then PC, PH and PSC should have identical distributions on target genes. Alternatively, it is possible that the association of PcG proteins is more dynamic: some proteins may have an important role at PREs, whilst others are required for localised spreading from the PRE.

The work in this chapter describes a comparison of the PC, PH and PSC distributions, concentrating on a different target from the BX-C, the *en-invected (inv)* genomic region. The characterised regulatory elements of *en* and *inv* are less extended than those of the BX-C genes, and are thus likely to be less complex. This allows the comparison of the PC distribution at a single locus with that of the BX-C. As a consequence of the smaller regulatory domains the analysis is also simplified.

The results shown below indicate that, as in the BX-C, PC is associated with discrete regulatory regions of the repressed genes of the *en-inv* locus which may correspond to PREs. In this case however, the PREs are located very close to the promoter region, probably reflecting the simpler organisation of the genes. Surprisingly, whilst PH and PSC are also associated with some PREs, they do not appear to be present at all the PC-binding loci, and are themselves associated with additional DNA elements. This suggests that the composition of the PcG protein complex may be variable. In addition, PH and PSC are associated with genes which are expressed in Schneider cells.

5.2. Expression analysis of genes in the *engrailed-invected* region

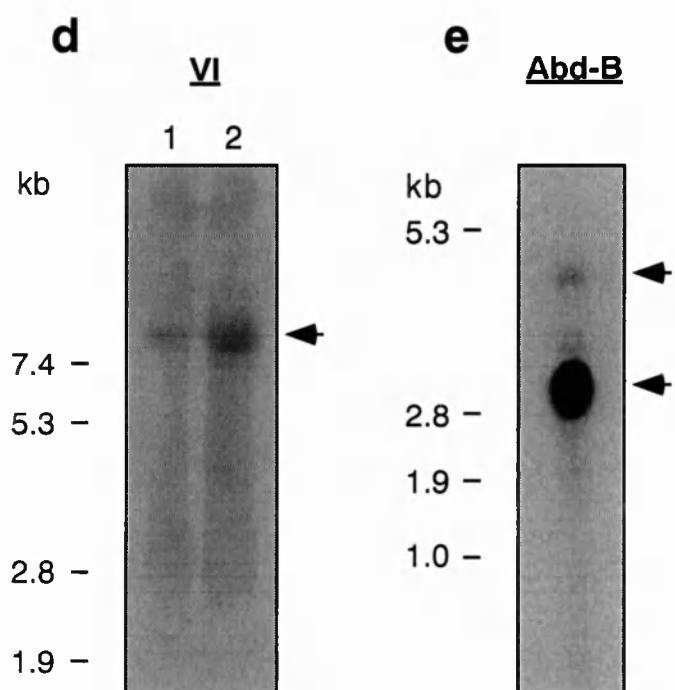
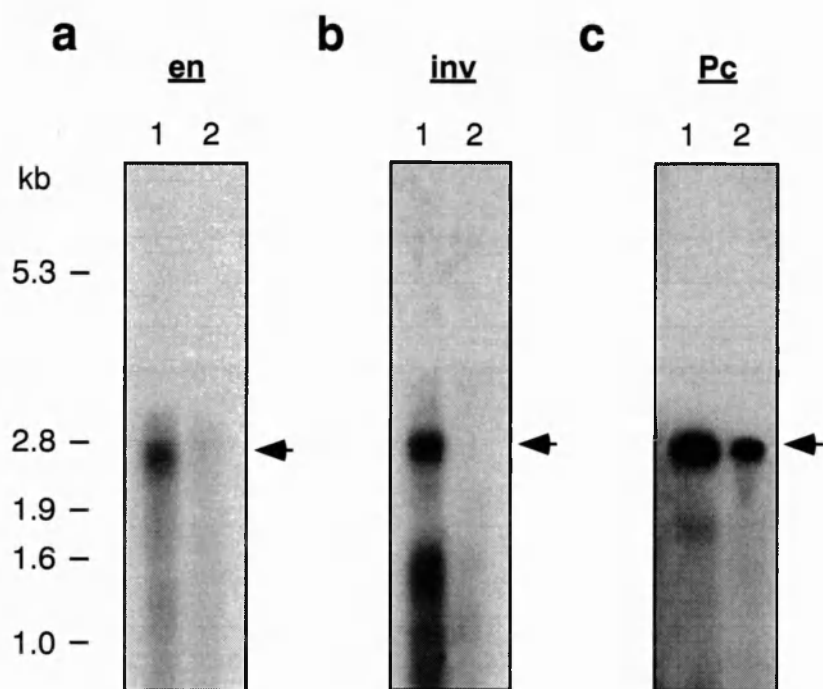
In addition to the selector gene *en*, two genes are present within the genomic walk chosen for the analysis (see figure A2). *inv* is located immediately downstream of *en*, with which it is highly homologous and encodes a redundant activity (Coleman et al., 1987;

Figure 5.1. Northern analysis of genes in the *engrailed-invected* genomic region and *Abdominal-B*

(a) - (c) Northern blot of 2 µg mRNA from 0 - 20 hour *Drosophila* embryos (lane 1) or from Schneider cells (lane 2), probed with DNA probes from the *en* cDNA (a), the first exon of *inv* (b) and, as a control, the *Pc* cDNA which is expressed in both embryos and Schneider cells (c). Arrows mark the positions of the major transcripts detected; *en* at 2.7 kb (Poole et al., 1985), *inv* at 2.7 kb (with a less abundant transcript at 1.2 kb)(Coleman et al., 1987) and *Pc* at 2.5 kb (Zink and Paro, 1989). *Pc* is expressed in both embryos and Schneider cells, whereas *en* and *inv* are only expressed in embryos.

(d) Northern blot of 10 µg total RNA from 0 - 20 hour *Drosophila* embryos (lane 1) or Schneider cells (lane 2), probed with a 3.6 kb BamHI fragment from λE13 which overlaps with transcript VI (see figure 5.4). An RNA of approximately 8.5 kb is detected (arrow), in agreement with that already reported (Drees et al., 1987). The same signal was detected with other probes from λE13.

(e) Northern blot of 2 µg mRNA from Schneider cells, hybridised with a 5.4 kb EcoRI fragment from λ8083, that contains sequences common to all *Abd-B* transcripts. A major transcript of approximately 3.5 kb is observed (lower arrow), which could therefore originate from either the B promoter or the γ promoter (Zavortink and Sakonju, 1989). As the γ promoter is covered by PcG proteins, it is assumed that the B promoter (free of PcG proteins) is active in Schneider cells. A low abundance transcript of 4.5 kb is also observed, which could represent transcription from the A promoter (Zavortink and Sakonju, 1989).



Tabata et al., 1995). Secondly, an 8.7 kb transcription unit of unknown function (gene VI) has been described upstream of *en* (Drees et al., 1987).

The expression state of *en*, *inv* and gene VI in Schneider cells was examined by Northern analysis. mRNA was prepared from Schneider cells and, as a control, from 0 - 20 hour embryos in which expression of these genes should be detectable. Equal amounts of embryonic and Schneider cell mRNA was separated by gel electrophoresis, and identical Northern filters were hybridised with probes either of an *en* cDNA or a genomic fragment including the first exon of *inv*. In both cases, a signal of the expected molecular weight could be detected in the lane containing embryonic RNA (figure 5.1a and b, lane 1), but not in the lanes carrying Schneider cell RNA (figure 5.1a and b, lane 2). As a control for the integrity of the Schneider cell RNA the filters were rehybridised with a probe against the *Pc* transcript, which should be present in both embryos and Schneider cells. This was clearly the case (figure 5.1c). As *en* and *inv* are easily detectable in RNA from embryos (in which only a proportion of the cells express the genes), but not in Schneider cells (in which the population is more homogeneous), it can be concluded that *en* and *inv* are not expressed in Schneider cells.

Gene VI has been reported to encode 8.7 and 8.9 kb poly (A)⁺ transcripts, which are expressed throughout embryogenesis (Drees et al., 1987). As expected, genomic probes from gene VI detect a transcript of approximately this size in embryonic RNA (figure 5.1d). Interestingly, this transcript is also present in Schneider cell RNA, indicating that gene VI is expressed in these cells. Therefore, as in the BX-C, it will be possible to analyse the association of the PcG proteins with a locus containing a mixture of active and inactive genes.

5.3. Association of Polycomb protein with the *engrailed-invected* walk

Schneider cells were cross-linked *in vivo* with formaldehyde, and soluble chromatin of an average size of 1 kb was generated by sonication. Chromatin associated with PC was immunoprecipitated with PC antibodies, and the DNA purified and amplified by linker-modified PCR. Amplified DNA from both PC and mock immunoprecipitations was used as a probe against a Southern of the *en-inv* genomic walk (figure 5.2a - c).

Control immunoprecipitated DNA gave low-level, approximately uniform hybridisation to most DNA fragments (figure 5.2b; hybridisation intensity is proportional to molecular weight). Restriction fragments containing repetitive elements (in λ clones E4, E7, E13 and E14) hybridise more strongly. Figure 5.3 shows the quantitation of the intensity of

Figure 5.2. Southern hybridisation of Polycomb group protein and GAGA factor immunoprecipitated DNA to the *engrailed-invected* genomic walk

(a) 1 µg DNA from 12 λ clones (E3 - E14)(Kuner et al., 1985) covering the *en-inv* genomic walk were digested with EcoRI and separated on a 0.5 % agarose gel, except for λE13 which was digested with EcoRI/BamHI. Lane M is a molecular weight marker of λ DNA digested with EcoRI-HindIII; the approximate positions of 5.0 kb, 2.0 kb and 1.0 kb are indicated to the left.

(b) Southern blot of the gel in (a) probed with DNA from a control immunoprecipitation without antibody. Note that hybridisation is approximately uniform, with signal intensity dependent on fragment size. Repetitive elements are in λE4 (a 1.9 kb EcoRI fragment), λE7 (a 3.4 kb EcoRI fragment), λE13 (a 2.1 kb BamHI fragment) and λE14 (a 4.7 kb EcoRI fragment).

(c) - (f) Identical blots probed with DNA from antibody immunoprecipitations; (c) PC; (d) GAGA factor; (e) PH; (f) PSC. With reference to figure 5.4, fragments strongly enriched by immunoprecipitations are: (c) PC, adjacent 3.5 kb and 1.5 kb EcoRI fragments of λE4 (A) and a 4.6 kb EcoRI fragment of λE8 (B); (d) GAGA factor, the 3.5 kb/λE4 and the 4.6 kb/λE8 fragments, and a 3.6 kb EcoRI/BamHI fragment from λE13; (e) PH, the 1.5 kb/λE4 and the 4.6 kb/λE8 fragments, and also two adjacent EcoRI fragments (2.8 kb and 2.2 kb) of λE12 (C); (f) PSC, the 4.6 kb/λE8 and the 2.8 kb/λE12 fragments. The 2.8 kb EcoRI fragment of λE12 overlaps with λE13 (1.5 kb and 1.0 kb EcoRI/BamHI fragments), but the hybridisation signal in this image is rather diffuse, due to the low percentage of the agarose gel.

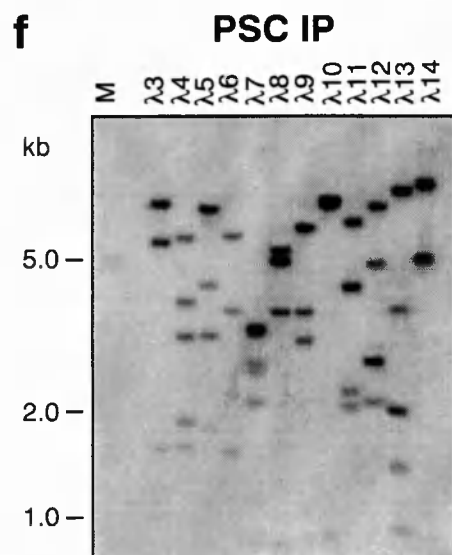
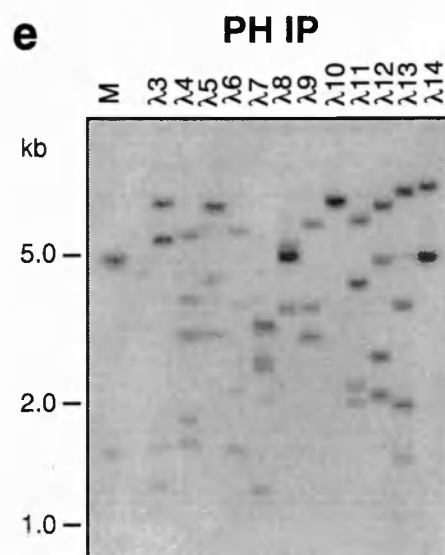
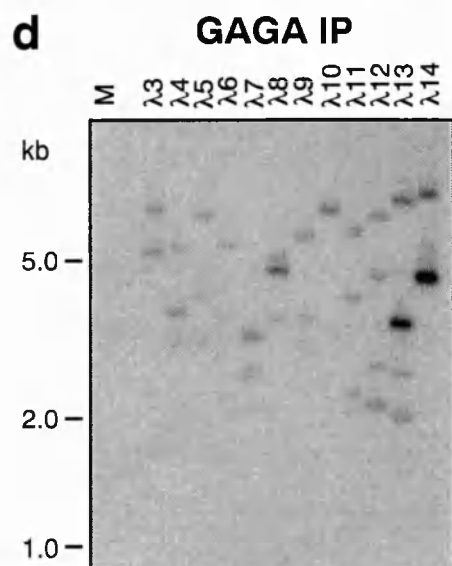
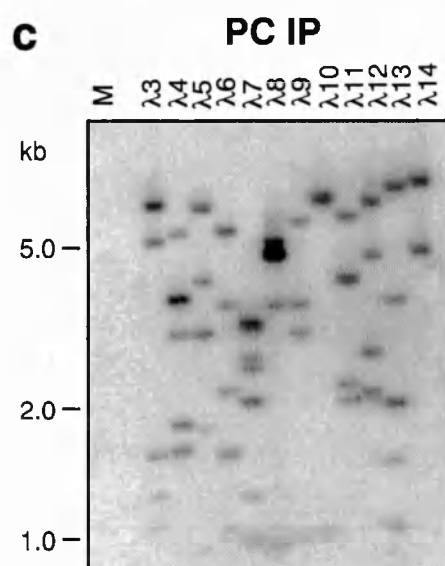
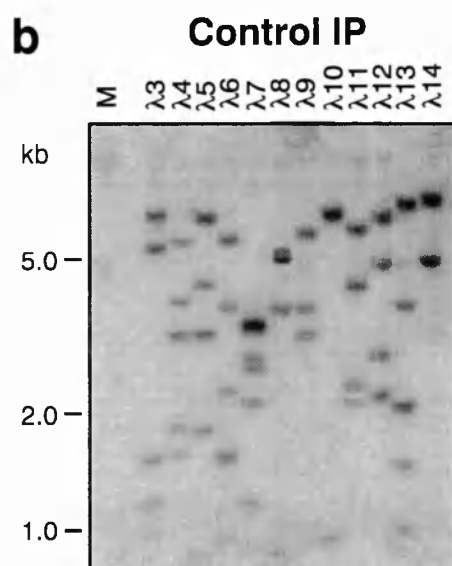
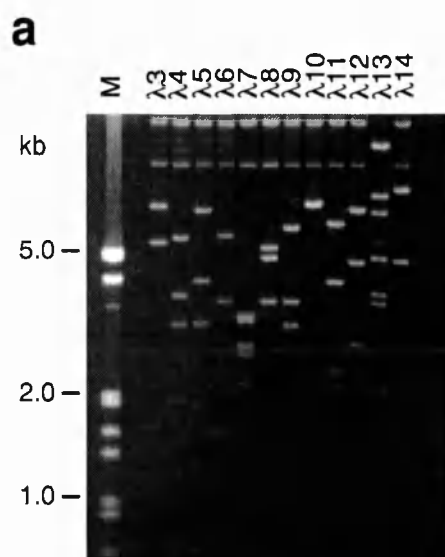
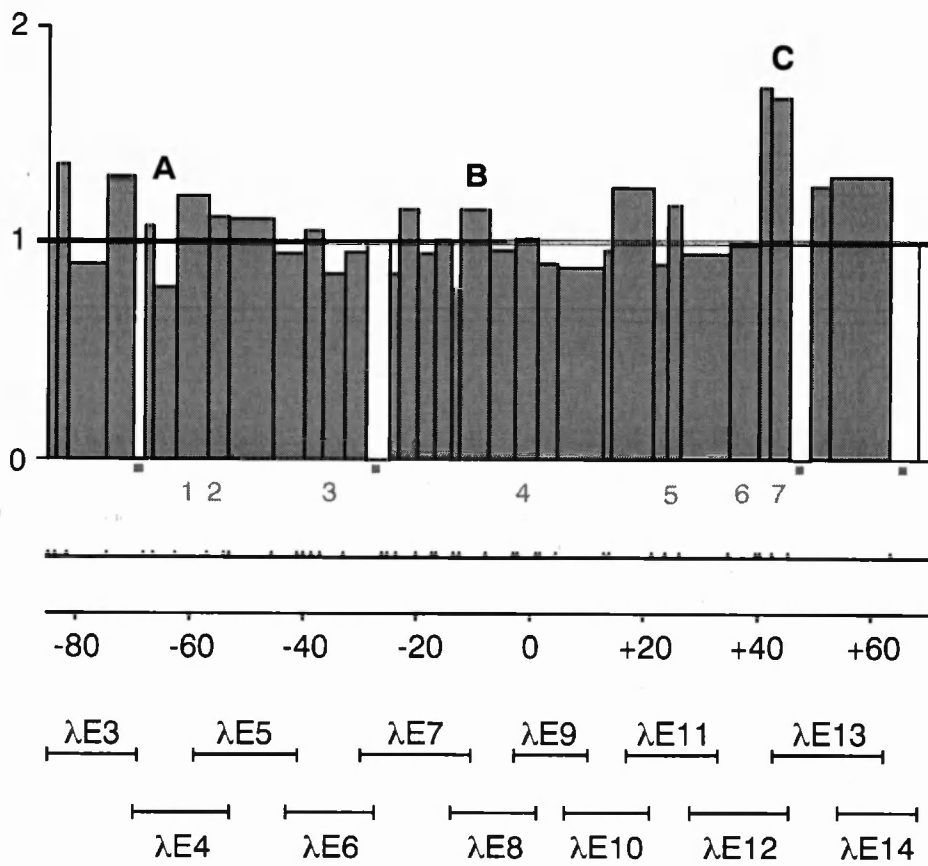


Figure 5.3. Hybridisation of control immunoprecipitated DNA to the *engrailed-invected* walk

The hybridisation signals in figure 5.2b were quantitated, corrected to account for molecular weight, and plotted on a map of the *en-inv* locus. Data was taken from at least two immunoprecipitation experiments from independent cross-linked chromatin preparations. The grey boxes represent relative signal intensity and the white boxes with a grey spot below are repetitive elements which could not be accurately quantitated. The mean hybridisation signal is depicted by the thick black line across the profile, and is arbitrarily set at 1.0. Above the figure the intron-exon structure of the *en* and *inv* genes is shown, together with the location of gene VI (the intron-exon structure is not known). At the top is shown the extent of lethal (dark grey bar) or non-lethal (light grey bar) breakpoint mutations which affect *en* function (Kuner et al., 1985). Underneath the figure are the positions of the λ phage (E3 - E14) used in the analysis, together with the restriction sites for EcoRI (E) and the map coordinates (Kuner et al., 1985). A detailed map of the locus is shown in Appendix A3. The fragments enriched most strongly by PcG protein immunoprecipitations are indicated by the letters A - C, and the positions of the fragments used for slot blot analysis are shown in grey (1 - 7).

Extent of en
breakpoint mutations



hybridisation of this control immunoprecipitated DNA to the *en-inv* walk. Interestingly, two restriction fragments in the vicinity of gene VI hybridise approximately 1.5-fold more strongly than the other fragments to control immunoprecipitated DNA, an effect not seen in genomic DNA hybridisations (data not shown). As gene VI is expressed, this difference is likely to be caused by the greater efficiency of sonication and thus greater amplification of DNA from active genes (see chapter 2.5).

Hybridisation of a PC immunoprecipitated DNA probe to the *en-inv* walk reveals that a number of DNA fragments are specifically enriched (figure 5.2c). Hybridisation signals were quantitated and plotted according to their position on the genomic map (figure 5.4a). The background level of hybridisation of the control immunoprecipitation (thick horizontal black line) was determined by slot blot analysis as before (table 5.1).

Table 5.1. Slot blot analysis of Polycomb group and trithorax group protein immunoprecipitated DNA in *engrailed-invected*

100 ng of control or antibody immunoprecipitated DNA was immobilised on nylon membrane by slot blot, and hybridised to probes 1 - 7 (all *EcoRI* fragments; see also figure 5.4). The enrichment in antibody immunoprecipitations (IP) is the mean of two experiments, and is calculated with respect to the hybridisation signal in control immunoprecipitations, normalised at 1.0 (n.d., no data).

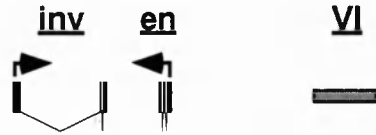
Fragment		Enrichment with respect to control IP			
		PC IP	GAGA IP	PH IP	PSC IP
1	5.8 kb (λ E4)	2.0	0.8	1.4	1.2
2	3.0 kb (λ E5)	1.1	n.d.	1.0	1.0
3	3.6 kb (λ E6)	1.2	1.1	0.9	0.9
4	3.7 kb (λ E8)	2.4	0.7	1.3	1.5
5	2.5 kb (λ E11)	1.5	n.d.	1.5	1.5
6	4.7 kb (λ E12)	1.0	0.8	1.2	1.3
7	2.8 kb (λ E12)	0.9	n.d.	2.5	2.5

Interestingly, the fragments most strongly enriched by PC-immunoprecipitations map to two regions; one upstream and overlapping the 5' end of the *en* gene, and the other near the 5' end of the *inv* gene. Association of PC with these fragments was mapped at higher resolution (figure 5.5). PC binding is maximal in a 1.0 kb element, 400 bp upstream of the *inv* start of transcription. However, in *en* PC associates with two distinct elements, one covering the first intron and the other 1 kb upstream from the start of transcription. Both these

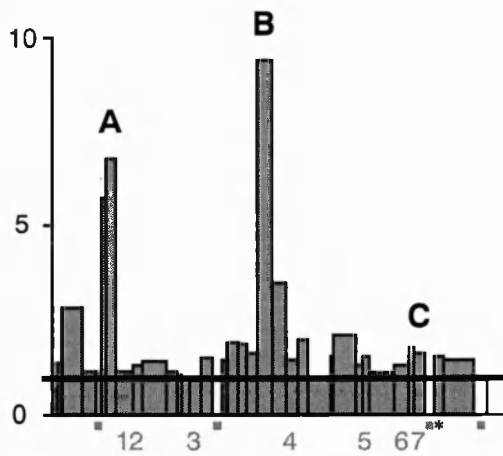
Figure 5.4. Distribution of Polycomb group proteins and GAGA factor on *engrailed-invected*

The hybridisation signals in panels (c) - (f) of figure 5.2 were quantitated, corrected to account for molecular weight, and plotted on a map of the *en-inv* locus; (a) PC; (b) GAGA factor; (c) PH; (d) PSC. Data was taken from at least two immunoprecipitation experiments from independent cross-linked chromatin preparations. The grey boxes represent relative signal intensity and the white boxes are repetitive elements which could not be accurately quantitated. The thick black line across each profile represents the approximate background level of hybridisation, as determined by comparison of plus and minus antibody immunoprecipitations by slot blot analysis (see table 5.1). Signals below this line are not considered to be enriched. The scale bar on the left indicates enrichment with respect to this background hybridisation, which is set at 1.0. Above panels (a) and (b) the intron-exon structure of the *en* and *inv* genes is shown, together with the location of gene VI (the intron-exon structure is not known). At the top is shown the extent of lethal (dark grey bar) or non-lethal (light grey bar) breakpoint mutations which affect *en* function (Kuner et al., 1985). Underneath figures (c) and (d) are the positions of the λ phage (E3 - E14) used in the analysis, together with the restriction sites for EcoRI (E) and the map coordinates (Kuner et al., 1985). A detailed map of the locus is shown in Appendix A3. The fragments enriched most strongly by PcG protein immunoprecipitations are indicated by the letters A - C, and the positions of the fragments used for slot blot analysis are shown in grey (1 - 7). The star denotes the position of the BamHI fragment used for Northern analysis.

Extent of en
breakpoint mutations



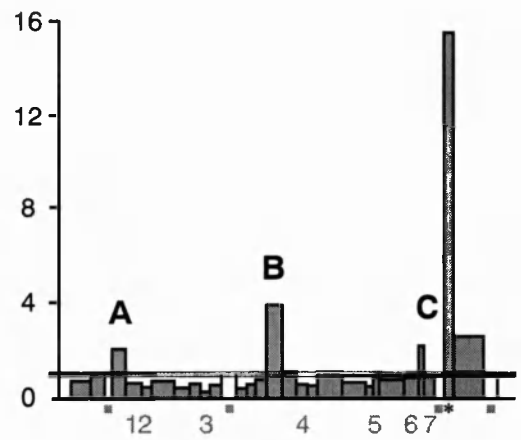
a **PC**



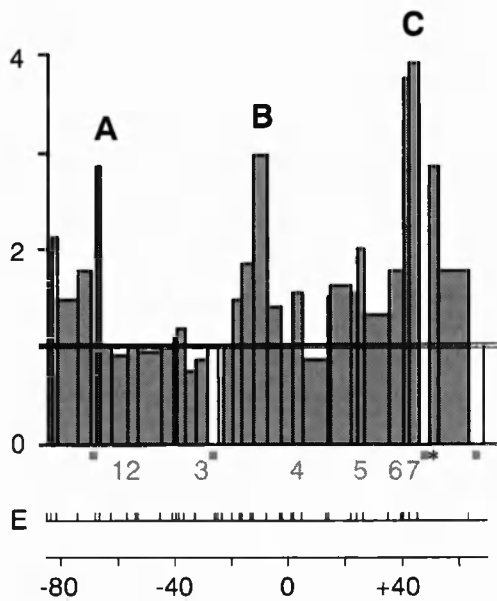
Extent of en
breakpoint mutations



b **GAGA**



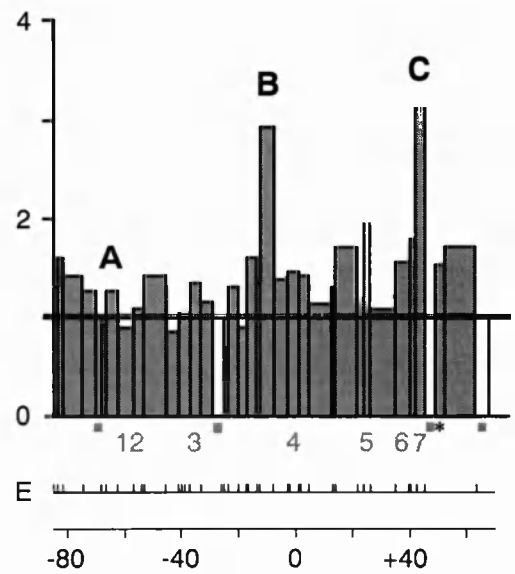
c **PH**



λ E3 λ E5 λ E7 λ E9 λ E11 λ E13

λ E4 λ E6 λ E8 λ E10 λ E12 λ E14

d **PSC**



λ E3 λ E5 λ E7 λ E9 λ E11 λ E13

λ E4 λ E6 λ E8 λ E10 λ E12 λ E14

regions have been implicated in regulation of *en* expression during embryogenesis (Hama et al., 1990; Kassis et al., 1991). Furthermore, the binding site upstream of *en* overlaps with a number of pairing sensitive elements which mediate PcG repression (Kassis, 1994).

Therefore, as in the BX-C, PC is associated with elements in inactive genes which have PRE activity, as discussed in chapter 3. In addition, the fact that GAGA factor is constitutively bound to PREs (chapter 3) could also be confirmed by analysis of the *en-inv* locus (figure 5.2d and 5.4b). Briefly, low levels of GAGA factor are found at both the PC binding sites in *en* (as seen by high resolution mapping; data not shown) and at *inv*. In addition, high levels of GAGA factor are associated with an element in the expressed gene *VI*, which may be comparable to the binding of GAGA factor to regulatory elements of *Abd-B*.

5.4. Mapping of Polyhomeotic and Posterior sex combs binding sites in the engrailed-invested region

Immunoprecipitations were carried out with PH and PSC antibodies, to investigate the distribution of PH and PSC on the *en* locus (figures 5.2e and f). The background level of hybridisation was determined (table 5.1). Although the enrichments relative to background are lower than was observed for PC (compare enrichments in figure 5.4), specific restriction fragments are reproducibly enriched. As both the PH and PSC antibodies can efficiently immunoprecipitate protein from nuclear extract, it is unlikely that the lower enrichments are caused by poor immunoprecipitation of chromatin. It may rather reflect the possibility that lower amounts of PH and PSC than PC are bound *in vivo*, or that PH and PSC are more distantly associated with the chromatin fibre than PC.

Quantitation of the hybridisation of PH and PSC immunoprecipitated DNA shows that PH and PSC, like PC, are associated with regulatory elements of the *en* gene (figure 5.4c and d). High resolution mapping shows that PH and PSC have an identical distribution to PC at both *en* elements, in the upstream region and the first intron (data not shown). The binding of all three proteins at *en* is consistent with the finding that they are members of the same multimeric complex, and suggests that this multimeric complex is functional at PREs.

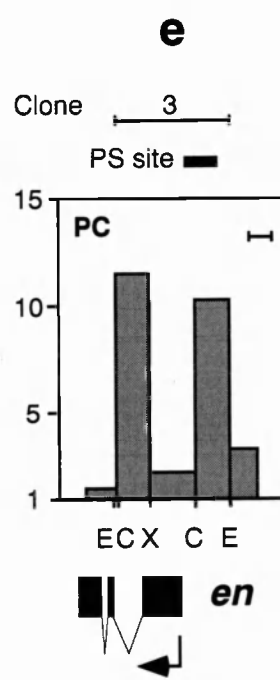
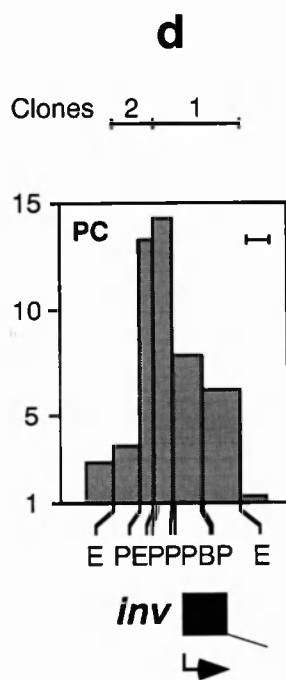
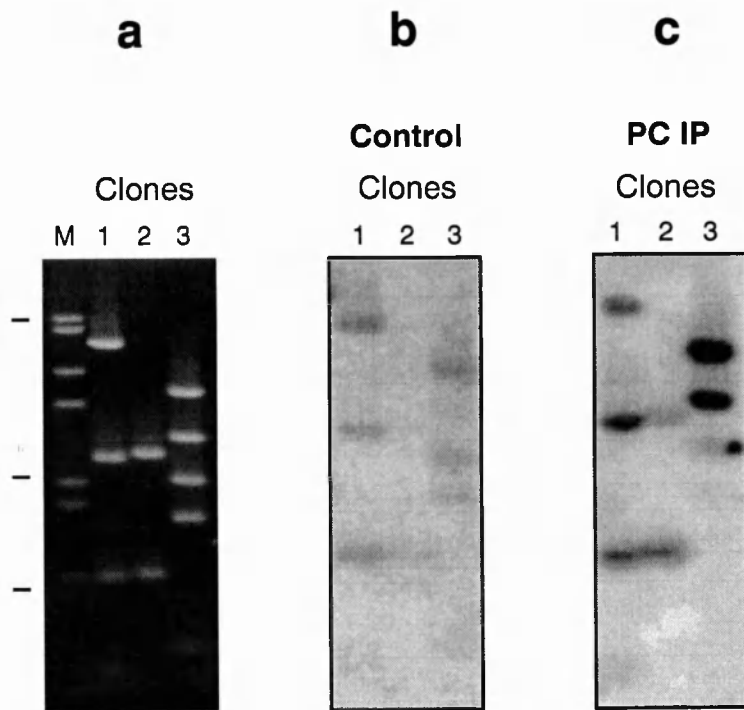
The distribution of PH and PSC on *inv* and gene *VI* is however not the same as PC. Firstly, the common PC/PH/PSC complex does not appear to function at *inv*: no PSC is associated with *inv*, and PH is associated with a much more restricted element than PC. This suggests that the PcG protein complex does not always have the same composition. To determine if this variability in composition is a common feature of the PcG protein complex,

Figure 5.5. High resolution mapping of proteins at putative PREs in *engrailed-invected*

(a) EcoRI fragments showing peak PC binding were subcloned in Bluescript and digested with various restriction enzymes before separating on an agarose gel. Clone 1, a 3.6 kb EcoRI fragment from λ E4 digested with BamHI-PstI; clone 2, an adjacent 1.5 kb EcoRI fragment from λ E4 digested with PstI-EcoRI; clone 3, a 4.6 kb EcoRI fragment from λ E8 digested with ClaI-XhoI-EcoRI. Lane M is a molecular weight marker of λ DNA digested with EcoRI-HindIII; the approximate positions of 2.0 kb, 1.0 kb and 0.5 kb are indicated to the left.

(b) and (c) Identical Southern filters from the gel in (a) probed with immunoprecipitated DNA; (b) control immunoprecipitation; (c) PC immunoprecipitation.

(d) and (e) Signals from Southern filter (c) were quantitated and plotted on a map of the genomic region. The height of the grey bars indicates relative hybridisation of DNA immunoprecipitated by PC to *inv* (d) or *en* (e), and the scale is the same as in figure 5.4a after subtraction of the background hybridisation (set at 1.0). The hybridisation intensity of the EcoRI fragments adjacent to the peak PC binding elements, as seen on figure 5.4a, are also shown for comparison. The scale bar at the top right is 1 kb. Underneath is shown the intron-exon structure of *en* or the first intron of *inv*. Restriction sites shown are BamHI (B), ClaI (C), EcoRI (E), PstI (P) and XhoI (X). Above is shown the extent of the DNA clones in panel (c), and the black bar is the position of the previously identified pairing-sensitive site (PS site) (Kassis, 1994).



the association of PH and PSC was examined with 7 elements in the BX-C which were previously shown to be strongly enriched by PC immunoprecipitations (see chapter 3; table 5.2). Interestingly, whilst some PREs (F and G) are associated with all three PcG proteins, the PcG complex at other PREs is either lacking PH (D and E), lacking PSC (B) or lacking both PH and PSC (A and C). The finding that not all PcG proteins are present at every PRE confirms that the PcG protein multimeric complex can have a variable composition at different target loci.

Table 5.2. Enrichment of Polycomb-associated PREs in the bithorax complex by Polyhomeotic and Posterior sex combs immunoprecipitations

EcoRI fragments (A - G) previously identified as peak PC binding sites (chapter 3) were analysed for enrichment in PH and PSC immunoprecipitations, by slot blot analysis and/or Southern analysis. The coordinates of the PREs are based on the complete sequence of the BX-C (Martin et al., 1995). An enrichment scored as positive is consistently greater than 2-fold.

¹ The 5316 bp *EcoRI* fragment immediately proximal to peak C is enriched in PSC immunoprecipitations

² The 4941 bp *EcoRI* fragment immediately proximal to peak E is equally enriched in PSC immunoprecipitations

PRE	BX-C coordinates	Immunoprecipitation		
		PC	PH	PSC
A	9636 - 12618	+	-	-
B (Mcp)	109688 - 115677	+	+	-
C (iab-4)	123772 - 131424	+	-	- ¹
D (iab-3)	152528 - 154422	+	-	+
E (iab-2)	159944 - 163814	+	-	+ ²
F (bx _d)	218241 - 221625	+	+	+
G (bx)	273301 - 275872	+	+	+
Fab-7	79681 - 88503	-	+	+

PH and PSC immunoprecipitations strongly enrich for fragments within the transcription unit VI. This is surprising, as the gene is expressed in Schneider cells, and is not associated with PC. Therefore, whereas PC binding is limited to regulatory elements of inactive genes, PH and PSC can associate with a transcribed locus independently of PC.

5.5. Regulation of gene VI expression by the Polycomb group proteins

The presence of PH and PSC within the gene VI transcription unit suggests that they may be involved in regulating the expression of this gene. For example, PH and PSC may form a functional complex which lacks PC and quantitatively down regulates (but does not completely silence) gene VI. Alternatively, the PH/PSC complex may not affect expression of gene VI in Schneider cells: PH and PSC could bind constitutively but be unable to establish silencing without additional PcG proteins such as PC. A final possibility is that in some contexts PH and PSC could be involved in transcriptional activation. To investigate the role of the PcG proteins in regulating gene VI, we carried out *in situ* hybridisation experiments in wild type and PcG mutant embryos, using a probe from the gene VI transcription unit.

In wild type embryos, gene VI is strongly and ubiquitously expressed throughout embryonic development (figure 5.6a - c). Surprisingly, the expression of gene VI is strongly down-regulated in homozygous *ph* mutant embryos (figure 5.6d and e). This suggests that PH is involved in positively regulating the expression of gene VI. Given the well-characterised role of PH as a repressor however, it is also possible that the effects of *ph* mutations on gene VI are indirect. Indeed, expression of gene VI is also down-regulated in *Pc* mutants, although not as dramatically as in *ph* mutants (figure 5.6f). As PC is not bound to sequences in gene VI, this result would support the idea that the apparent activating effect of PH is indirect, perhaps due to *ph* mutations causing ectopic expression of a direct repressor of gene VI. Therefore, the binding of PH and PSC to gene VI in Schneider cells is unlikely to represent an activation function.

5.6. Association of Polyhomeotic and Posterior sex combs with the Abdominal-B gene

The finding that PH and PSC are bound to the active gene VI suggests the possibility that the association of PH and PSC with their expressed target genes is a general phenomenon. The *Abd-B* gene is a suitable target gene to test this proposition, as it is expressed in Schneider cells and, like gene VI, is devoid of PC (see chapter 3). Western and Northern analyses indicate that in Schneider cells *Abd-B* is strongly expressed from the B promoter, and weakly expressed from the A promoter (Orlando and Paro, 1993; figure 5.1e). This suggests that Schneider cells are similar to cells in parasegment 14 of the embryo, in which expression of *Abd-B* is regulated by elements in the *iab-9* domain.

Figure 5.7 shows a comparison of the hybridisation of PC, PH and PSC immunoprecipitated DNA to *Abd-B*. As expected, the B promoter is devoid of all three PcG

Figure 5.6. Expression of gene VI in wild type and mutant embryos

Embryos were hybridised with digoxigenin-labelled DNA probes: a 2.0 kb EcoRI/BamHI fragment from λ E13, that recognises the 8.5 kb gene VI RNA, and/or the *Scr* cDNA clone pCY20 (Kuroiwa et al., 1985).

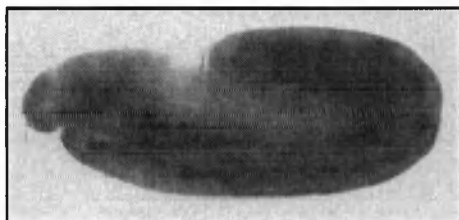
(a) - (c) Wild type embryos hybridised to a gene VI probe; (a) lateral view of a germ band extended embryo; (b) lateral view of an embryo which has undergone germ band retraction; (c) ventral view of a stage 16 embryo. Gene VI is strongly and ubiquitously expressed throughout most of embryogenesis, but fewer transcripts are detected in late embryonic stages.

(d) and (e) Stage 16 *ph* mutant embryos hybridised to a gene VI probe; (d) *Df(1)JA52*; (e) *Df(1)pn³⁸*. Late stage homozygous *ph* mutant embryos can be distinguished from wild type or heterozygotes by their severe abnormalities in the CNS (Smouse et al., 1988): these embryos show much reduced expression of gene VI.

(f) - (h) Analysis of gene VI expression in *Pc* mutants. *Scr* is used as a marker for *Pc* homozygous mutant embryos. (f) Stage 16 *Pc^{XT109}* mutant probed with *Scr* and gene VI probes. The *Scr* probe stains the anterior head (compare with (h)), and the remaining staining is due to the gene VI probe; note that the levels of gene VI expression are slightly lower than in wild type embryos. (g) Wild type stage 16 embryos, hybridised with an *Scr* probe. *Scr* is strongly expressed in the labial lobes in the anterior head, in the CNS and in the midgut (Kuroiwa et al., 1985). (h) Stage 16 *Pc^{XT109}* mutants, hybridised with an *Scr* probe. *Scr* expression is visible in the labial and maxillary lobes, but not in the CNS or the midgut (McKeon and Brock, 1991).

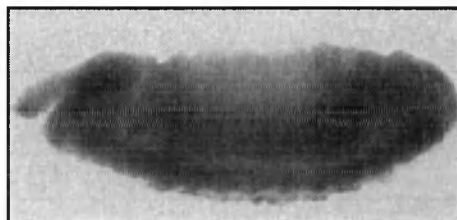
a

wt



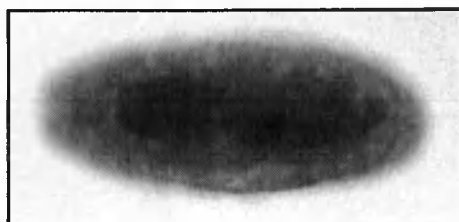
b

wt



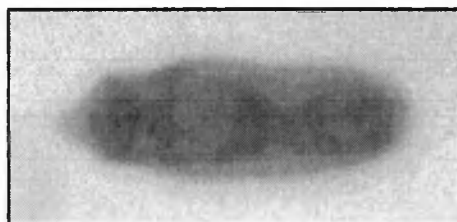
c

wt



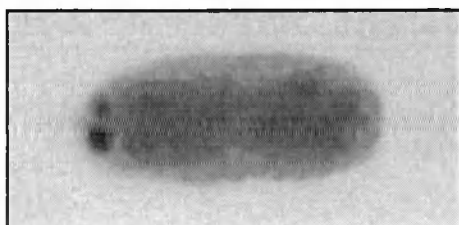
d

Df(1)JA52



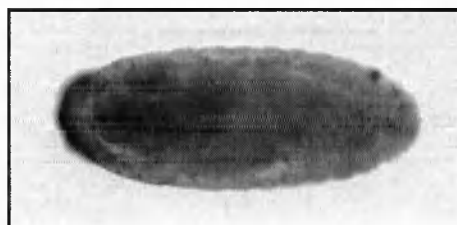
e

Df(1)pn³⁸



f

P_CXT109



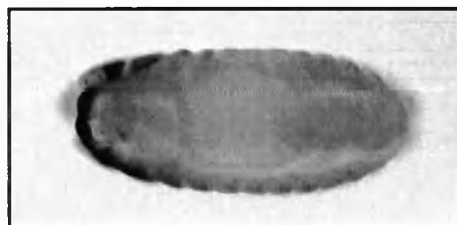
g

wt



h

P_CXT109



proteins, although the binding in the region of the A promoter cannot be determined due to an overlap with repetitive sequences. As described in table 5.2, PH and PSC immunoprecipitations do not enrich for the peak PC binding element A (overlapping the γ promoter). Other fragments in this vicinity are however associated with PH and/or PSC, and it may be that this regulatory region is unusually complex, and contains several PREs which regulate the different *Abd-B* promoters.

Both PH and PSC immunoprecipitations enrich for the *Fab-7* element (table 5.2) and a restriction fragment in the 3' region of *Abd-B* (X in figure 5.7g and h), which are relatively poorly enriched by PC. Both of these elements were previously shown to be strongly associated with GAGA factor, and element X, like *Fab-7*, is likely to be an *Abd-B* regulatory element (chapter 3). Interestingly, immunoprecipitations with PSC antibodies specifically enrich for additional restriction fragments in the first intron of *Abd-B*, which are contained within the active *iab-9* regulatory domain, and which must also be transcribed in Schneider cells (figure 5.7h). Similarly, the *empty spiracles (ems)* gene is expressed in Schneider cells (V. Orlando, unpublished data); again PSC is associated with an upstream fragment, covering a previously identified *ems* enhancer element (Jones and McGinnis, 1993). PC and PH are not found at this transcribed locus (data not shown).

5.7. Discussion

Polycomb binds to discrete elements in the regulatory regions of *engrailed* and *invected*

The binding of PC to the *en-inv* locus is consistent with its previously identified role in regulating the *en* gene (Busturia and Morata, 1988; Dura and Ingham, 1988; Moazed and O'Farrell, 1992). PC is associated with three discrete elements in the *en-inv* genomic region: one element upstream of the *inv* transcriptional start site, one element in the first intron of *en*, and finally an element 1 kb upstream of the *en* transcriptional start site.

Interestingly, both of the PC binding sites in *en* are in elements required for regulation of *en* expression in the embryo. Firstly, a 7.5 kb element immediately 5' to the *en* promoter directs striped expression of a *lacZ* reporter gene in the epidermis, and this pattern is maintained throughout embryonic development (Hama et al., 1990). Furthermore, expression is observed in many third instar larval tissues, and is restricted to the posterior compartments of imaginal discs. This confirms that the 7.5 kb region, which includes the PC binding site, is likely to contain sequences required for PcG regulation. In addition, a 1.6 kb region overlapping exactly with the upstream peak PC-binding site mediates pairing-sensitive repression of the *white* reporter gene in P-element transposon constructs (Kassis et

Figure 5.7. Association of Polycomb group proteins with *Abdominal-B* in Schneider cells

(a) 1 µg DNA from λ bacteriophage covering the *Abd-B* region was digested with EcoRI (lanes 1 and 2) or EcoRI/HindIII (lanes 3 - 5) and separated on a 0.6 % agarose gel; lane 1 (λ8083); lane 2 (λ8088); lane 3 (λ8095); lane 4 (λ8099); lane 5 (λ8106). DNA in lane 1 is not digested to completion. Lane M is a molecular weight marker, consisting of λ DNA digested with EcoRI/HindIII; the approximate positions of 5.0 kb, 3.5 kb, 2.0 kb and 1.0 kb are indicated adjacent to the marker.

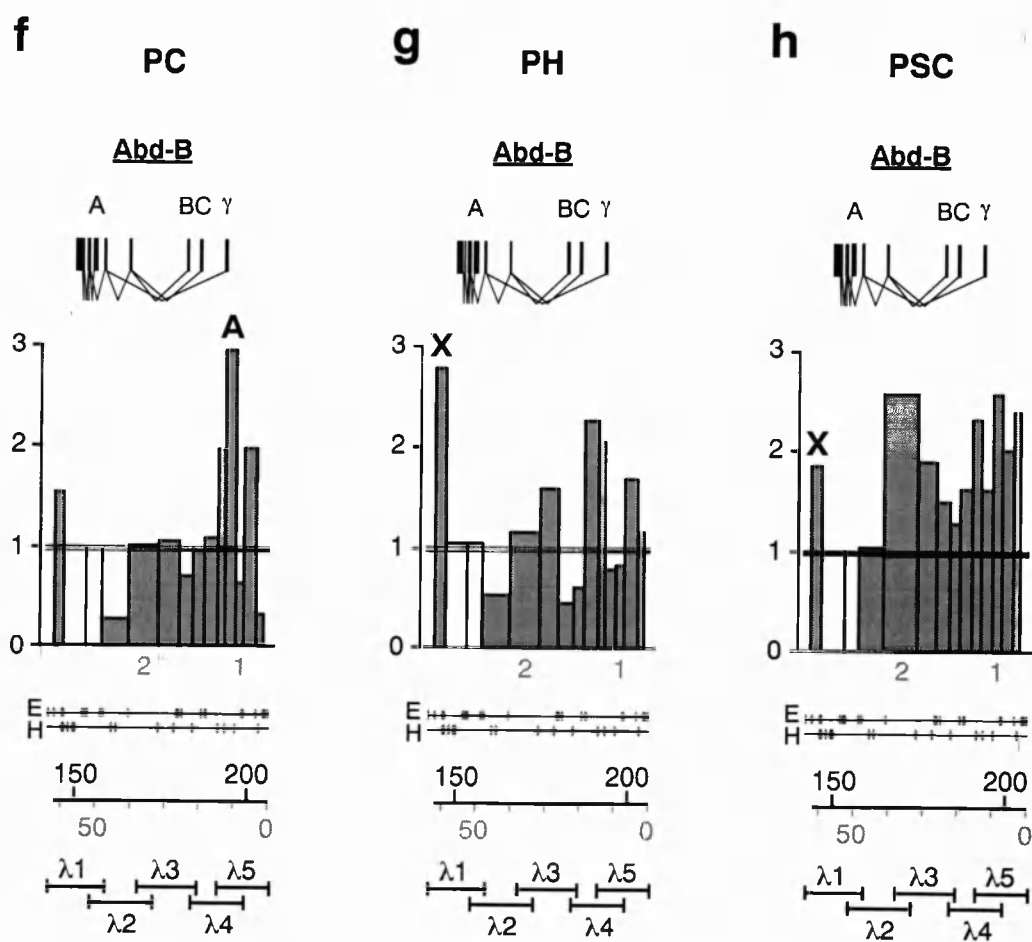
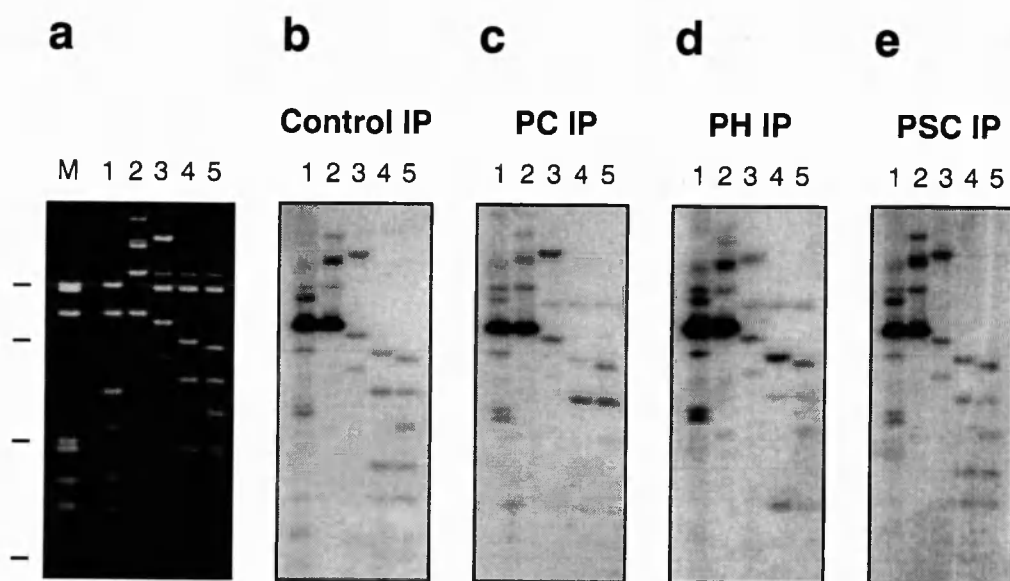
(b) - (e) Identical Southern filters from the gel in (a) probed with immunoprecipitated DNA.

(b) Control immunoprecipitation; note a strong repetitive element in lanes 1 and 2 (3.8 kb) and weaker repetitive elements in lanes 1 (5.4 kb) and 2 (7.0 kb). (c) PC

immunoprecipitation, see also figure 3.1; a 2.8 kb HindIII fragment in lanes 4 and 5 is enriched (marked as A in panel (f)). (d) PH immunoprecipitation; a 2.7 kb EcoRI fragment in lane 1 is enriched (marked as X in panels (g) and (h)). (e) PSC immunoprecipitations enrich for the large EcoRI fragment in lane 3.

(f) - (h) The hybridisation signals in (c) - (e) were quantitated, corrected to account for molecular weight, and plotted on a map of the *Abd-B* locus; (f) PC, see also figure 3.2a; (g) PH; (h) PSC. Data was taken from at least two immunoprecipitation experiments from independent cross-linked chromatin preparations. The grey boxes represent relative signal intensity and the white boxes are repetitive elements which could not be accurately quantitated. The thick black line across each profile represents the approximate background level of hybridisation, as determined by comparison of plus and minus antibody immunoprecipitations by slot blot analysis (using the fragments labelled 1 and 2 in grey). Signals below this line are not considered to be enriched. The scale bar on the left indicates enrichment with respect to this background hybridisation, which is arbitrarily set at 1.0.

Above the quantitations the intron-exon structure of the *Abd-B* gene is shown, together with the positions of the 4 promoters A, B, C and γ (Zavortink and Sakonju, 1989). Underneath are the positions of the λ phage used in the analysis and the restriction sites for EcoRI (E) and HindIII (H). The map coordinates shown in grey type are based on the complete sequence of the BX-C (Martin et al., 1995) and those in black type are the traditional coordinates (Karch et al., 1985).



al., 1991; Kassis, 1994). Mutations in PcG genes sometimes relieve repression, but this is dependent on the insertion site of the transposon. The strong binding of PC and other PcG proteins to this element supports the idea that it is a PRE which is required for the nucleation of a PcG protein complex.

Secondly, the first intron of *en* causes expression of a reporter gene in stripes during early development (Kassis, 1990), with expression fading later in development. This indicates that the intron contains early embryonic enhancers which initiate gene activity in posterior compartments, but does not contain late enhancers that maintain this expression. However, my results suggest that the intron may also contain a PRE which maintains gene silencing in anterior compartments.

The presence of a PRE in *inv* has not yet been reported. Indeed, it has been suggested that *en* and *inv* at least partially share regulatory DNA (Gustavson et al., 1996). Firstly, the two genes are expressed in essentially overlapping patterns. Secondly, a number of mutations in *en* regulatory sequences also affect *inv* expression. Finally, a breakpoint mutation which separates the two transcription units causes different aspects of the expression pattern to be allocated to either *en* or *inv*. However, the finding of strong PC binding near the *inv* transcriptional start site suggests that the maintenance of *inv* repression during development requires regulatory elements distinct from those of *en*. However, the absence of PSC binding at the *inv* PRE might also suggest that it is a "secondary" PRE which is only functional upon coordinate interaction with the "primary" PREs of *en*. As such, one could envisage that the *en* and *inv* PREs interact in the nucleus, resulting in a common cross-linked complex. Indeed, experiments with transgenes containing PREs appear to indicate that long-distance interactions between PREs can occur (Kassis, 1994; Pirrotta and Rastelli, 1994).

Polycomb group proteins form multiple different complexes

Despite being members of a common multimeric complex PC, PH and PSC do not always bind to identical target sites: PH and/or PSC are not found at the PC binding site at *inv*, nor at several PREs in the BX-C, whereas PC is absent from the complex at gene VI and *Abd-B*. This differential binding to a variety of target genes clearly indicates that although the PcG proteins interact together, the PcG protein complex must have a very flexible composition.

Previous studies have indicated that some PcG proteins may have functions which are independent of other PcG members. Firstly, the binding sites of PSC on polytene chromosomes are overlapping, but not identical to those of PC (Rastelli et al., 1993).

Secondly, *ph* mutants have a number of defects which are not normally associated with PcG mutants. In particular, there is extensive cell death in the ventral epidermis of *ph* mutants, a phenotype not normally found with other PcG alleles. Furthermore, there is an alteration in the pattern of axon pathways in the CNS, in which the wild type array of commissures and connectives is replaced by bundles of axons confined to the hemiganglia of origin (Dura et al., 1987; Smouse et al., 1988). This phenotype may be a result of abnormal expression of segmentation genes. Whilst defects in segmentation are common in PcG mutants, defects on such a scale in the nervous system are not. Given that the homeotic phenotype of *ph* mutants is not as strong as *Pc*, it is unlikely that a single PcG protein complex is involved in both processes. Finally, an extensive analysis of the genetic interactions between *ph* and other PcG mutants was carried out (Cheng et al., 1994). The range of phenotypes observed in embryos of double mutant combinations gave rise to the proposition that *ph* may perform different functions in conjunction with differing subsets of PcG proteins.

Are there a number of discrete PcG protein complexes, or is there a more dynamic interaction between the PcG proteins? It could be envisaged that a subset of PcG proteins (such as PC) form a "core" PcG protein complex at every silenced PRE, whereas the remaining PcG proteins may interact depending on chromosomal location but not be an integral part of the complex. Alternatively, the PcG protein complex may not be a discrete entity, but the members may be capable of multiple cooperative interactions with each other. The exact composition of the complex may rely both on the sequence of the PRE itself and on its chromosomal context. The more PcG proteins associating with the complex, the more stable the silencing. Conversely, mutations in any PcG gene (such as PC chromodomain mutations) may cause a general destabilisation of silencing, depending on the contribution of the mutant protein at a particular PRE.

Association of Polyhomeotic and Posterior sex combs with active genes

Intriguingly, whilst PC is exclusively associated with chromatin of inactive genes, PH and PSC are also bound to expressed genes, such as gene VI, *Abd-B* and *ems*. This could have a number of explanations: firstly the presence of PH and PSC at an expressed locus may reflect an additional activating role for these proteins. Whilst PcG proteins are by definition required for the maintenance of the inactive state of gene expression, there is a precedent for some of these proteins having other functions. In particular, the *E(z)* gene was initially characterised as a PcG gene, but a detailed analysis of the mutant phenotypes revealed that at some spatial and temporal stages of development *E(z)* acts as a *trxG* gene (LaJeunesse and

Shearn, 1996). Similarly, PSC was found to be exceptional when compared to other PcG members, in that it appeared to have an activating effect on *ph* expression (Fauvarque et al., 1995). However, it is unlikely that both PH and PSC have such an activating function. Although mutations in *ph* reduce gene VI transcription, mutations in *Pc* (not associated with gene VI) have the same effect. It is therefore more likely that the reduction in gene VI expression in PcG mutants is an indirect effect, caused by the derepression of a direct repressor. For example PcG mutations allow ectopic expression of the homeotic gene *Ultrabithorax* (*Ubx*). Later in development however levels of *Ubx* transcription are reduced compared to wild type, as the ectopically expressed *Abd-B* gene represses *Ubx* (Struhl and Akam, 1985; Wedeen et al., 1986).

As an alternative, it is possible that PH and PSC form a functional PcG protein complex that negatively regulates, but does not entirely inactivate gene VI. Indeed, these two PcG proteins might be part of a more general transcriptional mechanism which participates in modulated gene repression. For example, it was previously observed that PH negatively regulates its own expression (Fauvarque et al., 1995): this appears to involve the quantitative control of *ph* transcription rather than the maintenance of a fully inactive state as for the homeotic genes. In this case the establishment of full repression may require the addition of other PcG proteins to the complex, such as PC protein. For example, the strong binding of PH and PSC, but not PC, to regulatory elements in *iab-7* and *iab-8* (*Fab-7* and element X) of the BX-C may indicate that these domains are not fully inactivated, in contrast to the PC-bound proximal domains of the BX-C.

A final hypothesis is that some PcG proteins bind constitutively to target genes and thus act as markers for PREs. As PSC is associated with all three active genes which were examined (gene VI, *Abd-B* and *ems*), it seems likely that this protein at least may constitutively bind to many, if not all, of its target genes. As the trxB proteins have also been proposed to bind to target genes irrespective of transcriptional status (Chinwalla et al., 1995; Chang et al., 1995) it is tempting to speculate that constitutive binding of some PcG and trxB proteins is essential for correct functioning of both these classes of genes. Indeed, proteins bound constitutively to PREs may be essential components of a molecular switch that determines either an open or a repressed chromatin conformation.

- CHAPTER 6 -

Conclusions and outlook

6.1. The mechanism of silencing by Polycomb group proteins

A compaction model for the function of PcG proteins suggested that PcG proteins act in multimeric complexes that package inactive chromatin by spreading along the chromosome from an initiation point (the PRE) to a termination point (Locke et al., 1988; Paro, 1990). The results presented in this thesis show that in Schneider cells PC, PH and PSC are not homogeneously distributed over entire chromosomal domains of repressed target genes, but are highly enriched at PREs. As none of the three proteins examined is spread over extended domains, these results argue strongly against the involvement of the PcG proteins in general in chromatin compaction. The distribution of PcG proteins in the vicinity of PREs however suggests that PcG protein complexes spread locally over a distance of a few kilobases from the PRE.

How do these PcG protein complexes at PREs mediate silencing of target genes? One possibility is that the PcG proteins directly silence promoter function by looping and incorporating the promoter into the silencing complex (Bienz and Müller, 1995). The results presented here provide some support for such a model: in Schneider cells one of the peak PC binding sites overlaps with the *abd-A* promoter, and the *Antp* promoter is also enriched by PC immunoprecipitations. However the lack of strong PC binding at the *Ubx* and *en* promoters indicates that not all promoters appear to be directly associated with PC. An alternative hypothesis is that, although the PcG complexes do not spread to cover entire regulatory domains, localised spreading is sufficient to cover the important regulatory elements. For example, it has been shown that imaginal disc enhancers in the *bxd* domain of the *Ubx* gene are in close proximity (within a few kilobases) to the *bxd* PRE (Christen and Bienz, 1994; Pirrotta et al., 1995). Furthermore, the observation that many PREs are pairing sensitive (Kassis, 1994; Pirrotta and Rastelli, 1994) indicates that PcG protein complexes at PREs on homologous chromosomes associate together. It is therefore conceivable that PcG protein complexes at non-homologous sites also associate together within the nucleus: in addition to increasing the stability of silencing, this may create a structure in which enhancers are inaccessible to *trans*-acting factors.

It is not yet known how the silencing is stably maintained during the cell cycle, in particular throughout S phase and mitosis. Müller (1995) showed that sequences flanking a

PRE may be important for silencing to persist through many cell divisions. It is possible that spreading of PcG proteins into these flanking regions provides a mechanism for the transmission of the silenced state through S phase: PcG proteins ahead of the replication fork may induce the reassembly of the complex immediately behind the fork, without complete disruption of the complex. In mitosis however it was shown that only a small fraction of PC, PH and PSC remain on mitotic chromosomes (Buchenau et al., 1997). In this case, correct condensation of chromosomes may demand a reduction in the size of silencing complexes, but enough PcG proteins must remain to act as a marker for rebuilding the complete PcG protein complex when the chromosomes decondense.

How relevant are these results to the function of PcG proteins in the whole organism? The use of a cross-linking and immunoprecipitation method to analyse the distribution of proteins on target genes in the embryo is difficult, as the embryo contains many different cell types. However, preliminary results from this laboratory suggest that the distribution of PC on inactive genes is similar in Schneider cells and embryonic cells: immunoprecipitations against PC specifically enrich for PREs (V. Orlando and G. Cavalli, unpublished results).

Although the distribution of PC at PREs suggests a localised spreading of PcG protein complexes, it must be stressed that the results are only semi-quantitative, as there is some variation in the amplification of different DNA fragments (see figure 2.5). Therefore, a more rigorous analysis should be attempted in order to confirm that the apparent spreading is real. In yeast, the distribution of SIR proteins at telomeres was examined by amplifying DNA immunoprecipitated from cross-linked cells by quantitative PCR, using specific primer pairs at different distances from the telomere end (Hecht et al., 1996). A similar method would be suitable for analysing the distribution of PC in the vicinity of a particular PRE.

Finally, a number of peak PC binding DNA fragments were identified, notably in *iab-4*, *iab-9* and *inv*, which have not yet been characterised as PREs. The coincidence of known PREs and PC binding sites strongly suggests that these elements are in fact also PREs, and that immunoprecipitation of cross-linked chromatin may be used to identify DNA fragments with PRE activity. To confirm this, these identified elements could be tested in reporter constructs in transgenic flies: as PREs they should be able to reproduce the boundaries of the expression domains of homeotic genes in a PcG gene-dependent manner and recruit PcG proteins to the insertion site of the transposon.

6.2. The role of GAGA factor at distant regulatory elements

GAGA factor is most well characterised as a protein that binds to promoter regions, where it is thought to contribute to gene activation by creating an "open" chromatin conformation. In the BX-C of Schneider cells however, GAGA factor is bound to PREs, regulatory elements that are distant from the promoter. This is consistent with other evidence that GAGA factor acts as a trxG protein. Indeed, preliminary results from this laboratory indicate that other trxG proteins such as TRX are also bound to PREs (V. Orlando, unpublished results). This data indicates that the main sites of activity of TRX and GAGA factor are the sites at which PcG proteins bind, and suggests that the trxG proteins may antagonise the PcG proteins via a direct interaction at upstream regulatory elements. PcG proteins and TRX/GAGA factor bound to a particular PRE may thus compete at some level to control the state of activity of the entire regulatory domain. However, PREs may not be the only sites at which trxG proteins operate. GAGA factor may for example have an additional function at the promoters of active trxG regulated genes (although it is not apparent at the *Abd-B* promoters in Schneider cells), and it is quite possible that other trxG gene products antagonise PcG silencing by utilising diverse mechanisms. It is not yet clear if the low level of GAGA factor binding at PC-bound PREs indicates a role for GAGA factor in PcG function. It may be possible to address such a question using PRE-containing reporter constructs in transgenic animals. For example, the effects on PcG silencing of mutating GAGA factor binding sites could be examined.

In addition, GAGA factor is strongly associated with the *Fab-7* boundary element. By comparison with the presumed function of enhancers of PEV, it is possible that GAGA factor has a role in preventing the inappropriate spreading of inactive chromatin from the *Fab-7* PRE into the *iab-6* regulatory domains. Such a hypothesis could be tested by analysis of reporter gene constructs containing combinations of the *Fab-7* boundary element and PRE in *Trl* mutant backgrounds. In addition, the formaldehyde cross-linking and immunoprecipitation technique, applied to embryos, may allow the spreading of PcG complexes on either side of the *Fab-7* PRE to be mapped, in the presence or absence of the boundary element.

6.3. Multiple forms of Polycomb group protein complexes

The results presented here show that PcG protein complexes at different chromosomal locations do not all have the same complement of PcG proteins. Such a finding is consistent with the variability in the mutant phenotypes of PcG genes, but leaves many unanswered

questions. Firstly, are there a number of discrete PcG protein complexes, or is the composition of the "complex" dynamic? Secondly, do different PcG protein complexes lead to a quantitative or qualitative difference in silencing? Finally, how is the formation of the PcG protein complexes differentially regulated at PREs?

Our understanding of the PcG protein complex may be aided by a more systematic investigation of the interaction of PcG proteins with isolated PREs in reporter constructs in flies. For example, the silencing ability of most characterised PREs (with the exception of *Scr*) has not been tested in several different PcG mutant backgrounds. Interestingly, the two PREs of *Scr* respond differently to mutations in a variety of PcG genes (Gindhart and Kaufmann, 1995), although the lack of an effect on silencing does not necessarily imply that the mutated PcG protein has no role at that PRE. Nevertheless, studies of this type with other PREs may be rewarding. In addition, it may be possible to analyse the interaction of cloned PcG gene products with PREs more directly, both at the level of polytene chromosome binding and by analysing isolated PREs in formaldehyde cross-linked chromatin. It would also be interesting to investigate if the sequences flanking the PRE affect the composition of the complex: for example reporter constructs may be compared which contain either a minimal PRE or a PRE with 5 - 10 kb of flanking region. Such studies should yield important insights into the structure and function of the PcG silencing complex.

- CHAPTER 7 -

Materials and Methods

7.1. Materials

Fly stocks

The wild type flies used in this study were Canton S, and all mutant alleles are described in "The genome of *Drosophila melanogaster*" (Lindsley and Zimm, 1992). *Pc^{XT109}* is an X-ray induced null allele of *Pc*, kept as a balanced stock over *TM3 Ser. Df(1)JA52* and *Df(1)pn³⁸* are X-ray induced deficiencies that uncover *ph*, and were kept as balanced stocks over FM6. Transgenic flies expressing 1-266 PC-lacZ and 1-266 PC Δ 42-65-lacZ are described in Messmer and Paro (1992), and flies expressing 1-390 PC-lacZ were generated by Sabine Messmer.

Tissue culture cell lines

The *Drosophila melanogaster* Schneider cell line SL2 (Schneider, 1972) was grown at 25 °C in Schneider's *Drosophila* medium (Serva), supplemented with 100 U/ml penicillin, 100 µg/ml streptomycin (Gibco) and 10 % heat-inactivated fetal calf serum (Boehringer). Cell densities were maintained between 1.5×10^6 and 6×10^6 cells/ml, and were diluted every 2 - 3 days.

Antibodies

PC antibodies are affinity purified rabbit polyclonals, directed against a PC- β -galactosidase fusion protein (containing the carboxy-terminal 199 amino acids of PC) as previously described (Zink and Paro, 1989). Antibodies recognising bacterially-expressed GAGA factor were kindly provided by Peter Becker (EMBL, Heidelberg), and the PH and PSC antibodies are described in chapter 4.

Bacterial strains

The following bacterial strains were used:

(i) *E. coli* LE392

hsd R574, (*rK*⁻, *mK*⁺), *supE44*, *supF58*, *lacY1* or $\Delta(lacIZY)6$, *galK2*, *galT22*, *metB1*, *trpR55*

Used for propagating λ bacteriophage. Contains the amber suppressors *supE* and *supF*, and is therefore permissive for λ vectors such as Charon 4 that carry amber mutations.

(ii) *E. coli* XL1-Blue (Stratagene)

recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac(F' proAB, lacI^q, ZΔM15, Tn10 (Tet^r))

Used for blue-white screening of plasmids, and single strand DNA rescue of phagemid DNA. Contains a selectable (tetracycline resistant) F' episome, carrying the lacI^q gene, which produces 10-fold more lac repressor than the wild type gene.

(iii) *E. coli* BL21 (DE3)

F⁻, ompT, hsdS_B, (r_B⁻, m_B⁻), dcm, gal, λ{DE3}

Used for protein expression. The lysogenic λ phage encodes T7 RNA polymerase under the control of the *lacUV5* promoter. Expression is induced by IPTG.

(iv) *E. coli* M15 [pREP4] (Qiagen)

Nal^S, Str^S, rif^S, lac⁻, ara⁻, gal⁻, mlt^r, F⁻, recA⁺, uvr⁺

Used for protein expression. Contains the multi-copy plasmid pREP4, which carries the *lacI* gene (encoding the lac repressor).

Plasmid vectors

(i) pQE-30, -31 and -32 expression vectors (Qiagen)

Contains the bacteriophage T5 promoter and two lac operator sites upstream of a translation start site, a sequence encoding a 6xhistidine affinity tag, a polylinker and a translation stop site. The three vector variants contain the polylinker in all three reading frames. Expression is repressed in strains carrying high amounts of lac repressor, and is induced by IPTG.

(ii) pRSETA. B and C expression vectors (Invitrogen)

Contains the bacteriophage T7 RNA polymerase promoter upstream of a translation start site, a sequence encoding a 6xhistidine affinity tag, a polylinker and a translation stop site. The three vector variants contain the polylinker in all three reading frames. Expression occurs if the bacteriophage T7 RNA polymerase gene is induced from a host strain.

Plasmid subclones

Fragments showing peak PC or GAGA factor binding were subcloned into the corresponding restriction enzyme site in the polylinker of Bluescript KS+ (Stratagene) for further analysis (tables 7.1 and 7.2). In addition, a 4 kb EcoRI/KpnI fragment surrounding the *Antennapedia* P1 promoter shows PRE activity (Zink et al., 1991). 8 overlapping subclones of this fragment are present in Bluescript (Stratagene; see table 7.3).

Table 7.1. Subclones of the bithorax complex

The coordinates of the BX-C subclones are based on the complete published sequence (SEQ89E; accession number U31961) of the bithorax complex (Martin et al., 1995).

Subclone	BX-C fragment	Coordinates
PC Peak A (p8106.1)	2982 bp HindIII fragment	9636 - 12618
PC Peak B (p5989)	5989 bp EcoRI fragment	109688 - 115677
PC Peak C (p7652)	7652 bp EcoRI fragment	123772 - 131424
PC Peak D (p1894)	1894 bp EcoRI fragment	152528 - 154422
PC Peak E (p3870)	3870 bp EcoRI fragment	159944 - 163814
PC Peak F (p3384)	3384 bp EcoRI fragment	218241 - 221625
PC Peak G (p2571)	2571 bp EcoRI fragment	273301 - 275872
Fab-7 clone 1 (p4344)	4344 bp EcoRI fragment	79681 - 84025
Fab-7 clone 2 (p4389)	4389 bp EcoRI fragment	84114 - 88503

Table 7.2. Subclones of the engrailed-inverted walk

Subclone	DNA fragment
pE4.1	3.6 kb EcoRI fragment from λE4
pE4.2	1.5 kb EcoRI fragment from λE4
pE8.1	4.6 kb EcoRI fragment from λE8

Table 7.3. Subclones of the Antennapedia P1 promoter region

	<i>Antp</i> fragment	coordinates	vector cloning site	re-isolate insert
BSX1	EcoRI - HpaI	1 - 616	EcoRI - AccI	EcoRI - XhoI
BSX2	HhaI - HindII	551 - 1096	SmaI	EcoRI - BamHI
BSX3	HinfI - BglII	1048 - 1605	SmaI	EcoRI - BamHI
BSX4	AflII - XhoI	1439 - 2075	SmaI	EcoRI - BamHI
BSX5	HinfI - HinfI	2041 - 2482	SmaI	EcoRI - BamHI
BSX6	AhaII - MscI	2481 - 3136	SmaI	EcoRI - XbaI
BSX7	HhaI - HhaI	3038 - 3678	SmaI	EcoRI - BamHI
BSX8	NsiI - KpnI	3392 - 3968	KpnI - PstI	KpnI - BamHI

Bacteriophage P1 clones

DNA was isolated from 5 bacteriophage P1 clones covering 280 kb of the BX-C (Martin et al., 1995). The P1 clone spanning the remaining 60 kb grew very poorly and could not be used. A single colony of bacteria containing P1 bacteriophage was inoculated into 5 ml LB medium, supplemented with 25 µg/ml kanamycin, and grown overnight at 37 °C. 1 ml of this overnight culture was added to a flask containing 500 ml LB and 25 µg/ml kanamycin, and grown at 37 °C until the O.D.₆₀₀ was 0.3 - 0.5 (approximately 2 - 3 hours). High copy number replication of the bacteriophage was induced by addition of IPTG to 1 mM. The cultures were allowed to grow for a further 2 - 3 hours, until the O.D.₆₀₀ was 1.3 - 1.5. Bacteria were harvested by centrifugation and bacteriophage DNA purified using standard Qiagen maxiprep columns, as instructed by the manufacturer.

Bacteriophage λ clones

DNA was isolated from 12 bacteriophage λ clones (in Charon 4A vectors) covering the *en-inv* locus (Kuner et al., 1985) and 5 clones (in Charon 4 vectors) from the *Abd-B* region of the BX-C (Karch et al., 1985). Bacteriophage were amplified in *E.coli* strain LE392 using standard methods (Sambrook et al., 1989), and λ DNA was purified using Qiagen DNA purification columns as recommended by the manufacturer.

7.2. Immunoprecipitation of *in vivo* formaldehyde cross-linked chromatin from tissue culture cells

This method is based on a previously-described formaldehyde cross-linking and immunoprecipitation method (Orlando and Paro, 1993; Orlando et al., 1997).

In vivo formaldehyde cross-linking and purification of chromatin

1 µCi/ml (methyl-³H)-thymidine was added to 10⁹ log-phase SL2 cells, and the cells were grown at 25 °C for 36 - 48 hours. One tenth volume of fixation solution was added (11 % formaldehyde, 100 mM NaCl, 50 mM Tris.HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA) and the cells fixed for 10 minutes at 25 °C followed by 50 minutes at 4 °C. Fixation was stopped by addition of glycine to 125 mM. Cells were centrifuged at 1000 g for 10 minutes at 4 °C, washed with ice-cold PBS and again centrifuged. The pellet was resuspended in 20 ml wash solution A (10 mM Tris.HCl pH 8.0, 10 mM EDTA, 0.5 mM EGTA, 0.25 % Triton X-100), incubated for 10 minutes at room temperature, centrifuged, and resuspended in 20 ml wash solution B (200 mM NaCl, 10 mM Tris.HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA). After a

further 10 minute incubation at room temperature, cells were again centrifuged and resuspended in 12 ml sonication buffer (10 mM Tris.HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA).

Sonication was carried out in 6 ml aliquots using four 30 second bursts of a Branson Model 250 sonifier microtip at maximum setting. Samples were adjusted to 0.5 % with sarkosyl and mixed for 10 minutes at room temperature, before centrifuging at 15000 g for 15 minutes at room temperature to remove debris. Caesium chloride was added to a density of 1.42 g/ml, and the samples made up to 20 ml for centrifugation in a Beckmann SW55Ti rotor at 40000 rpm for 72 hours at 20 °C.

0.4 ml fractions were collected from the bottom of the gradient, and DNA-chromatin containing fractions (peak ³H fractions) were dialysed overnight into 5 % glycerol, 10 mM Tris.HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA. Samples were frozen in 500 µl aliquots (containing 30 - 60 µg DNA) at -70 °C.

Immunoprecipitation from cross-linked chromatin

500 µl aliquots of chromatin were adjusted to RIPA buffer by sequential addition of Triton X-100 to 1 %, sodium deoxycholate to 0.1 %, SDS to 0.1 %, NaCl to 140 mM and PMSF to 1 mM. 30 µl of 50 % v/v protein A Sepharose CL4B (Sigma), pre-equilibrated in RIPA buffer, was added, incubated for 1 hour at 4 °C, and removed by centrifugation at 15000 g for 20 seconds. 2 - 5 µg antibody was incubated with the chromatin sample overnight at 4 °C. Specific antibody-chromatin complexes were isolated by adding 30 µl Protein A Sepharose for 3 hours at 4 °C. The Protein A Sepharose suspension was then washed five times for ten minutes each in RIPA buffer, once in LiCl buffer (250 mM LiCl, 0.5 % NP-40, 0.5 % sodium deoxycholate, 10 mM Tris.HCl pH 8.0, 1 mM EDTA) and twice in TE buffer (10 mM Tris.HCl pH 8.0, 1 mM EDTA), centrifuging for 20 seconds between each wash. Samples were finally resuspended in 100 µl TE buffer.

Purification of immunoprecipitated DNA

Cross-links were reversed by incubating in 50 µg/ml RNase A for 30 minutes at 37 °C. Samples were adjusted to 0.5 % SDS, 500 µg/ml proteinase K, and incubated overnight at 37 °C, followed by 6 hours at 65 °C. Samples were phenol chloroform extracted, chloroform extracted and ethanol precipitated in the presence of 20 µg glycogen as carrier.

Amplification of immunoprecipitated DNA using linker-modified PCR

The immunoprecipitation strategy yields approximately 1 ng DNA from starting material of 50 µg. Therefore, to have enough DNA to be used as a probe on a Southern blot, the DNA was amplified using linker-modified PCR. In early experiments immunoprecipitated DNA was first digested with a restriction enzyme, and ligated to an appropriate linker with a compatible end (section (ii) below). Oligonucleotide primers homologous to the linker DNA were subsequently used to amplify the internal genomic sequences. However amplification was not uniform between restriction fragments of different size and an alternative strategy was developed, in which a blunt-ended linker was ligated directly to the sonicated, immunoprecipitated DNA without prior restriction (section (iii) below). This method amplified DNA approximately linearly (see chapter 2).

(i) Preparation of linker DNA

1 nmol oligonucleotide DNA was phosphorylated using T4 polynucleotide kinase (Boehringer), in a 20 µl reaction with 1 x corresponding buffer, 500 µM ATP and 20 U enzyme. The reaction was incubated at 37 °C for 60 minutes. Excess nucleotide was removed by precipitation. The sample was diluted to 60 µl in dH₂O, and 240 µl 5 M ammonium acetate and 750 µl ice-cold 100 % ethanol were added. After a 30 minute incubation on ice, the sample was centrifuged at 4 °C for 20 minutes at 15000 g. The pellet was washed in 70 % ethanol, recentrifuged, air-dried, and resuspended in TE buffer at a concentration of 10 µM.

Linkers were made by annealing two complementary oligonucleotides. Equal molar quantities of oligonucleotides were mixed and incubated at 70 °C for 5 minutes followed by the appropriate annealing temperature for 5 minutes. Samples were then allowed to cool slowly to room temperature, and stored at -20 °C.

(ii) Restriction digestion and ligation of linker to immunoprecipitated DNA

Immunoprecipitated DNA was resuspended in 9 µl NdeII buffer (10 mM Tris.HCl pH 7.6, 150 mM NaCl, 10 mM MgCl₂) and digested for 1 hour with 4 U NdeII (Boehringer). DNA was ethanol precipitated using 20 µg glycogen as carrier, and was resuspended in 9 µl ligation buffer (10 mM Tris.HCl pH 7.5, 10 mM MgCl₂, 2 mM ATP, 5 mM β-mercaptoethanol), containing linker at a final concentration of 0.5 µM. Linkers were made by annealing a 24-mer oligonucleotide of sequence 5'-GAT CAG AAG CTT GAA TTC GAG CAG and a 20-mer of sequence 5'-CTG CTC GAA TTC AAG CTT CT, of which only the 24-mer is phosphorylated at the 5' end.

(iii) Direct ligation of linker to immunoprecipitated DNA

Approximately one third of the immunoprecipitated DNA (1 ng) was resuspended in 7 μ l dH₂O, to which 1 μ l 10x ligation buffer (0.5 M Tris.HCl pH 7.6, 125 mM MgCl₂, 250 mM DTT, 12.5 mM ATP) was added, and linkers to a final concentration of 0.1 μ M. Linkers were made by annealing a 24-mer oligonucleotide of sequence 5'-AGA AGC TTG AAT TCG AGC AGT CAG and a 20-mer of sequence 5'-CTG CTC GAA TTC AAG CTT CT, of which only the 24-mer is phosphorylated at the 5' end.

(iv) Ligation and amplification of DNA

Ligation was carried out at 4 °C overnight, after the addition of 4 U T4 DNA ligase. Amplification was carried out directly in a 100 μ l reaction using Taq polymerase (Boehringer) and the corresponding buffer, deoxynucleotides at 250 μ M, and 1 μ M primer (the 20-mer oligonucleotide used above). The amplification cycles were 1 cycle of 94 °C for 2 minutes; 35 cycles of 94 °C for 1 minute, 55 °C for 1 minute, 72 °C for 3 minutes; and 1 cycle of 94 °C for 1 minute, 55 °C for 1 minute, 72 °C for 10 minutes.

Amplified DNA was phenol-chloroform extracted and ethanol precipitated, and linker DNA sequences removed by HindIII digestion. Amplified DNA was subsequently purified from linker DNA using Qiagen PCR purification columns, according to the manufacturer's conditions. 100 ng of this purified DNA was either analysed by slot blot, or 50 - 100 ng used as a hybridisation probe on Southern blots.

Slot blot analysis of immunoprecipitated DNA

To determine the relative enrichments of potential target sequences in immunoprecipitated chromatin, DNA from either plus antibody or control (no antibody) immunoprecipitations was immobilised on Genescreen plus nylon membrane (NEN) by slot blot. 100 ng immunoprecipitated DNA was diluted to 10 μ l in dH₂O, and denatured by the addition of an equal volume of 0.5 M NaOH, 1 M NaCl for 10 minutes at room temperature. Samples were then diluted on ice to 200 μ l in 0.1 x SSC, 125 mM NaOH. Genescreen plus membrane was prewetted in 0.4 M Tris.HCl pH 7.5, and the slot blot apparatus (Schleicher and Schuell) assembled according to the manufacturer's instructions. Samples were loaded into the manifold, and after 30 minutes suction was applied. When all the fluid had entered the wells, the membrane was removed and neutralised for 2 minutes in 0.5 M NaCl, 0.5 M Tris.HCl pH 7.5. DNA was fixed to the membrane by baking at 80 °C for 2 hours. Membranes were hybridised to potential target DNA probes as described for Southern blots.

Southern blot analysis of immunoprecipitated DNA

DNA of potential target sequences was restricted and separated by agarose gel electrophoresis, before blotting onto Genescreen plus membrane by standard methods (Sambrook et al., 1989). Hybridisation was essentially as described in Church and Gilbert (1984). Membranes were prehybridised for 2 hours at 65 °C in 0.5 M sodium phosphate buffer pH 7.2, 7 % SDS, 1 mM EDTA, 1 % BSA. 50 - 100 ng DNA was labelled with ($\alpha^{32}\text{P}$)-dATP using a Boehringer random priming kit, denatured, and added to the prehybridisation solution. Hybridisation was carried out overnight at 65 °C. Filters were washed once for 10 minutes at 65 °C in 40 mM sodium phosphate buffer pH 7.2, 5 % SDS, 0.5 % BSA, 1 mM EDTA, and three times for 5 minutes at 65 °C in 40 mM sodium phosphate buffer pH 7.2, 1 % SDS, 1 mM EDTA.

Quantitation of signal on slot blots and Southern blots

Filters were exposed overnight to a Phosphorimager screen, and scanned using the Molecular Dynamics PhosphorImager apparatus. The integrated value of each band was determined using PC software (ImageQuant) connected to the Phosphorimager.

For Southern analysis of genomic walks the resulting values were normalised to take account of the molecular weight difference between the bands, and the values plotted along the genomic walk.

7.3. Purification of Polyhomeotic and Posterior sex combs fusion proteins

Cloning of expression constructs

Expression constructs were made by cloning appropriate regions of the *ph* and *Psc* cDNAs (figure 4.1)(Brunk et al., 1991; DeCamillis et al., 1992; van Lohuizen et al., 1991) into either pQE vectors (Qiagen) or pRSET vectors (Invitrogen):

(i) pOEph5

1304 bp XhoI-SalI fragment (bases 750 - 2054 of published cDNA sequence), into the SalI site of pQE-31.

(ii) pRSETph2

675 bp PstI fragment (bases 2770 - 3445 of cDNA sequence), into the PstI site of pRSETC.

(iii) pRSETpsc7

1145 bp PstI-HindIII fragment (bases 1421 - 2566 of published cDNA sequence), into the PstI-HindIII sites of pRSETB.

(iv) pRSETpsc8

1044 bp HindIII fragment (bases 2566 - 3610 of cDNA sequence), into the HindIII site of pRSETB.

pQE and pRSET vectors contain a coding sequence for 6 histidine residues; downstream is a multicloning site in all three reading frames (pQE-30, pQE-31, pQE-32; and pRSETA, pRSETB, pRSETC), such that the cDNA can be inserted in frame with the six-histidine residues.

Expression from the pQE vectors is under the control of the phage T5 promoter, which contains two lac operator sites. Constructs are transformed into an *E.coli* strain M15 (Qiagen), which contains multiple copies of the kanamycin-resistant plasmid pREP4. High levels of lac repressor are produced from this plasmid, thus preventing protein expression. Addition of IPTG inactivates the repressor, and the fusion protein is expressed.

pRSET vectors express fusion protein under the control of the phage T7 promoter. For cloning, constructs are transformed into XLI-Blue *E.coli* (Stratagene) which do not contain the T7 RNA polymerase. For expression, constructs are transformed into *E. coli* strain BL21 (DE3) (Studier and Moffatt, 1986), in which the T7 RNA polymerase gene is induced after addition of IPTG.

Expression and purification of fusion proteins

For pQE vectors in *E.coli* strain M15(pREP4), a single colony from a freshly-streaked plate was inoculated into 50 ml LB medium, supplemented with 100 µg/ml ampicillin and 25 µg/ml kanamycin, and grown overnight, shaking, at 37 °C. 1 ml of this culture was inoculated into 500 ml fresh medium plus antibiotics, and grown until the O.D.₆₀₀ was 0.5.

pRSET vectors in *E.coli* strain BL21(DE3) were streaked from a glycerol stock onto an LB agar plate containing 100 µg/ml ampicillin, and grown overnight at 37 °C. All the colonies were scraped off a single plate, resuspended in 500 ml LB plus ampicillin, and grown until the O.D.₆₀₀ was 0.6 - 0.8.

Cultures were induced by the addition of IPTG to 2 mM, and the culture grown for a further 2 hours at 37 °C. Cells were harvested by spinning at 4000 g for 20 minutes, and the cell pellet was frozen at -70 °C.

Fusion protein was purified under denaturing conditions, essentially as recommended by Qiagen. The cell pellet was thawed on ice for 15 minutes, and resuspended in 10 ml buffer A (6 M GuHCl, 0.1 M NaH₂PO₄, 10 mM Tris.HCl pH 8.0). After stirring for one hour at room temperature, debris was removed by centrifugation at 10000 g for 15

minutes at 4 °C. The supernatant was applied to a pre-equilibrated 4 ml nickel chelate column (Chelating Sepharose, fast flow; Pharmacia Biotech.), at a flow rate of 0.4 ml per minute. The column was washed with 40 ml buffer A, 20 ml buffer B (as buffer A, but 8 M urea instead of GuHCl) and finally with 80 - 100 ml buffer C (as buffer B, but pH 6.3).

Protein was eluted with 30 ml buffer C, containing 250 mM imidazole (pH 6.3). 1.5 ml fractions were collected and the fractions containing fusion protein were identified by analysing 10 µl on 10 % SDS polyacrylamide gels followed by Coomassie staining (Sambrook et al., 1989). Protein was then dialysed stepwise to remove urea, first overnight at 4 °C into 100 mM Tris.HCl pH 7.5, 150 mM NaCl, 5 mM MgCl₂, 0.005 % Tween-20, 4 M urea. Dialysis was continued for 5 hours into the same buffer containing 2 M urea, for 5 hours into buffer with 1 M urea, and finally into buffer without urea. The protein concentration was measured using the Biorad protein assay reagent, and protein was then stored at -70 °C.

Charging and storage of nickel chelate columns

Before use, 4 ml nickel chelate columns were washed sequentially with 20 ml dH₂O, 10 ml 100 mM NiSO₄, 10 ml dH₂O and 20 ml buffer A. After protein purification, residual protein was removed by washing in 20 ml 6 M GuHCl, 0.2 M acetic acid, followed by 20 ml 0.1 M EDTA. Columns were finally washed in 20 ml dH₂O and equilibrated to 70 % ethanol, before storing at 4 °C.

7.4. Purification of polyclonal antibodies against Polyhomeotic and Posterior sex combs proteins

Immunisation of rabbits

Antibodies were produced in rabbits by standard methods (Harlow and Lane, 1988). Immediately after taking 10 ml preimmune serum, 250 µg fusion protein and 250 µl RAS (R-adjuvans system; Pan Systems) were injected. After three weeks the rabbits were boosted with the same mixture; and then every four weeks thereafter. 30 ml blood was collected one week after every boost, and the serum tested on Western blots containing either fusion protein or nuclear extracts from Schneider cells.

Preparation of affinity columns

For each fusion protein an affinity column for purifying polyclonal antibodies was prepared, by covalently coupling fusion protein to an activated support. Fusion protein, prepared as described in section 7.3, was dialysed overnight into 100 mM MOPS pH 7.5,

150 mM NaCl, 5 mM MgCl₂ and concentrated in a Centricon 30 device (Amicon) until the concentration was 2 - 5 mg/ml. 2 ml Affigel 10 (Biorad) was washed sequentially with 10 ml isopropanol, 10 ml dH₂O and three times in 10 ml 100 mM MOPS pH 7.5. 2 - 10 mg fusion protein was stirred with the gel for 4 hours at 4 °C. The gel was washed three times in 10 ml MOPS pH 7.5, and then incubated for 1 hour at 4 °C in 100 mM ethanolamine. Finally the gel was washed in 10 ml glycine.HCl pH 2.5, two times in 0.1 M Tris.HCl pH 7.5, 0.5 M NaCl, and four times in PBS, before storage at 4 °C in PBS plus 0.02 % sodium azide. Efficiency of coupling was determined by measuring the protein concentration before and after incubation with the Affigel.

Purification of polyclonal antibodies

Serum from immunised rabbits was collected, and allowed to clot by incubating at 37 °C for 1 hour, and then at 4 °C overnight. Serum was centrifuged at 4000 g for 15 minutes, followed by 10000 g for 15 minutes, and the supernatant stored at 4 °C in the presence of 0.02 % sodium azide. Immunoglobulins were first purified on a 5 ml Affigel protein A column (Biorad). The column was equilibrated in PBS, washed with 20 ml 0.1 M glycine pH 2.8, and again in PBS until the pH was neutral. 10 ml serum was diluted with one volume PBS, and applied to the column two times at a flow rate of 1 ml per minute. The column was washed with PBS until the O.D.₂₈₀ of the flow-through was less than 0.01.

Immunoglobulins were eluted with 10 ml 0.1 M glycine pH 2.8, and twenty 1 ml fractions were collected into tubes containing 50 µl Tris.HCl pH 8.0. Immunoglobulin-containing fractions were identified by measuring the O.D.₂₈₀. Peak fractions were combined, adjusted to 1 x PBS and applied at least three times to the affinity column, preequilibrated with PBS. Washing and elution was the same as for the protein A column. Peak antibody fractions were concentrated by Speed-vac, adjusted to 1 x PBS and dialysed overnight into PBS. They were finally stored at 4 °C in the presence of 1 % BSA and 0.02 % sodium azide.

7.5. Immunoprecipitation from nuclear extracts

Preparation of nuclear extracts from *Drosophila melanogaster* embryos and tissue culture cells

(i) Embryonic nuclear extract

Nuclear extract was prepared essentially as in Franke et al. (1992). 2 g *Drosophila* embryos were collected 0 - 20 hours after egg laying and washed with 0.4 % NaCl, 0.03 % Triton X-100. After dechorionation for 2 - 3 minutes in 3 % sodium hypochlorite, embryos

were again washed thoroughly in the NaCl/Triton solution, allowed briefly to dry, and frozen in liquid nitrogen.

Frozen embryos were resuspended in 8 ml buffer B (15 mM Hepes pH 7.6, 10 mM KCl, 5 mM MgCl₂, 0.1 mM EDTA, 0.5 mM EGTA, 1 mM PMSF, 1 mM DTT, 2 µg/ml leupeptin, 2 µg/ml pepstatin, 2 µg/ml aprotinin), and stored on ice until thawed. They were then homogenised with 15 strokes in a Kontex glass homogeniser with a type A pestle, and filtered through two layers of Miracloth (Calbiochem). Nuclei were pelleted by centrifugation at 2000 g for 10 minutes at 4 °C, and resuspended in 4 ml buffer B. To further purify the nuclei, 2 ml aliquots of this suspension were layered onto a 2 ml cushion of buffer B + 0.8 M sucrose in centrifuge tubes, and spun at 8000 rpm for 10 minutes at 4 °C, in a Beckman SW60Ti rotor.

Nuclei were then resuspended in 4 ml buffer B containing 150 mM KCl. 440 µl 4 M ammonium sulphate pH 7.6 was added dropwise, and after a 15 minute incubation on ice the sample was spun at 31000 rpm for 1 hour in a Beckman SW60Ti rotor. The pellet was discarded and 0.3 g/ml ammonium sulphate was slowly added to the supernatant before centrifugation at 15000 g for 15 minutes at 4 °C. The protein pellet was resuspended in 300 µl buffer C (25 mM Hepes pH 7.6, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5 mM PMSF, 10 % glycerol). The nuclear extract was frozen in liquid nitrogen and stored at -80 °C, and the concentration was usually 10-15 mg/ml.

(ii) Nuclear extract from Schneider cells

2 x 10⁹ log-phase SL2 cells were centrifuged at 1000 g for 5 minutes and the cell pellet resuspended in 8 ml buffer B as above. The procedure was then identical to that for embryonic nuclear extract, except that the sucrose gradient step was omitted.

Immunoprecipitation

25 µl (250 µg) nuclear extract was adjusted to 250 µl with RIPA buffer (140 mM NaCl, 0.1 % SDS, 1 % Triton X-100, 0.1 % sodium deoxycholate, 10 mM Tris.HCl pH 8.0, 1 mM EDTA, 1 mM PMSF, 1 mM DTT, 2 µg/ml leupeptin, 2 µg/ml pepstatin, 2 µg/ml aprotinin), and an appropriate volume of antibody was added. Samples were incubated, rotating, for 1 hour at 4 °C; 25 µl 50 % v/v Protein A Sepharose CL4B (Sigma; preequilibrated in RIPA buffer) was added and the incubation continued for a further hour at 4°C. The protein A Sepharose was centrifuged at 15000 g for 30 seconds, and the pellet washed five times for 2 minutes each in RIPA buffer, one time in LiCl buffer (0.25 M LiCl, 0.5 % NP-40, 0.5 % sodium deoxycholate, 10 mM Tris.HCl pH 8.0, 1 mM EDTA, 1 mM PMSF, 1 mM DTT), and

finally two times in TE buffer (10 mM Tris.HCl pH 8.0, 1 mM EDTA, 1 mM PMSF, 1 mM DTT). The Protein A Sepharose was again centrifuged and resuspended in 25 µl protein gel sample buffer (100 mM Tris.HCl pH 6.8, 4 % SDS, 20 % glycerol, 0.2 M DTT, 0.2 % bromophenol blue), before analysis on SDS polyacrylamide gels.

SDS polyacrylamide gel electrophoresis and Western blotting

All analysis of proteins by SDS polyacrylamide gel electrophoresis and Western blotting was by standard methods (Sambrook et al., 1989). Western blots were analysed by using either alkaline phosphatase antibody detection or ECL (Amersham), as instructed by the manufacturer.

7.6. Northern blot analysis of gene expression

Total RNA preparation from *Drosophila melanogaster* embryos or tissue culture cells

RNA was prepared from 0 - 20 hour embryos, which had been dechorionated and stored at -70 °C (section 7.5), and from SL2 cells which had been centrifuged at 1000 g for 5 minutes, washed once in PBS, and the cell pellet stored at -70 °C. RNA was prepared by extraction with guanidinium thiocyanate and centrifugation in caesium chloride (Sambrook et al., 1989).

Preparation of mRNA

Poly (A)⁺ RNA was purified by one round of affinity chromatography on oligo(dT)-cellulose (Boehringer), using standard methods (Sambrook et al., 1989).

Denaturing gel electrophoresis of RNA

(i) Preparation of agarose gels

1 - 1.2 % agarose was dissolved in 115 ml 1 x MOPS buffer (10 x MOPS buffer is 0.2 M MOPS, 50 mM NaAc, 10 mM EDTA, pH adjusted to 7.0 with NaOH), and allowed to cool. 5 ml 37 % formaldehyde was added, immediately before pouring into the gel chamber (in a fume cupboard). Gels were run in 1 x MOPS buffer, at 3 - 4 V/cm at 4 °C, until the bromophenol blue dye was 2/3 - 3/4 down the gel. To visualise the RNA under UV, gels were stained for 5 minutes in a solution of 5 µg/ml ethidium bromide in dH₂O, and destained for 2 hours in dH₂O.

(ii) Preparation of RNA sample for electrophoresis

2 µg mRNA or 20 µg total RNA was resuspended in 5 - 10 µl dH₂O. 2 volumes denaturation cocktail was added (2 : 1 : 10; 10 x MOPS buffer : 37 % formaldehyde : formamide) and samples were incubated at 70 °C for 10 minutes, before cooling on ice. 0.2 volumes 6 x loading buffer was added (0.5 % SDS, 25 % glycerol, 25 mM EDTA, bromophenol blue, xylene cyanol), before loading on the denaturing gel

Northern blotting

RNA from formaldehyde gels was transferred directly by capillary blot to Genescreen plus nylon membrane (NEN), overnight in 10 x SSC. The filter was rinsed in dH₂O and UV cross-linked in a Stratalinker (Stratagene) according to the manufacturers instructions. Finally, filters were baked at 80 °C for 2 hours to reverse any remaining formaldehyde cross-links.

Hybridisation of Northern blots

Hybridisation was carried out as previously described for Southern blots (section 7.2)(Church and Gilbert, 1984), with DNA probes which had been labelled with ($\alpha^{32}\text{P}$)-dATP (NEN) using a random priming kit (Boehringer) as recommended. After washing, blots were exposed to X-ray film in the presence of intensifier screens.

7.7. *In situ* hybridisation to embryos

This method is based on the "*In situ* hybridisation to embryos with non-radioactive probes" protocol of D. Tautz (Ashburner, 1989).

Fixation of embryos

0 - 16 hour embryos were collected and washed with a solution of 0.4 % NaCl, 0.03 % Triton X-100, before dechorionating for 2.5 minutes in 3 % sodium hypochlorite. Embryos were again washed thoroughly with the NaCl/Triton X-100 solution, and with dH₂O. Embryos were then transferred to a scintillation vial containing 2.0 ml fixation solution (0.4 ml 20% paraformaldehyde and 1.6 ml HEM buffer: 100 mM Hepes pH 6.9, 2 mM MgSO₄, 1 mM EGTA). 8 ml heptane was added, and the embryos fixed by shaking vigorously for 15 - 20 minutes.

The phases were allowed to separate with the embryos at the interface, and the lower aqueous phase was discarded. To remove the vitelline membrane 10 ml of methanol was

added and the embryos shaken for 10 seconds. Most of the upper heptane phase was then removed, and replaced with methanol. After shaking again for 10 seconds, all of the embryos were devitellinised and had sunk to the bottom of the tube. The embryos were transferred with a Pasteur pipette into an Eppendorf tube, and were washed in ME solution (50 mM EGTA, 90 % methanol). Embryos were then rehydrated through a series of ME: PP (4 % paraformaldehyde in PBS) washes: 5 minutes in 1 ml 70 % ME, 30 % PP; 5 minutes in 1 ml 50 % ME: 50 % PP; 5 minutes in 1 ml 30 % ME, 70 % PP and finally 20 minutes in 1 ml PP.

Pretreatment of embryos

All the following incubation steps were carried out in a 1 ml volume, on a rotating wheel at room temperature unless otherwise indicated.

Embryos were washed three times for 5 minutes each in PBST (0.3 % Tween-20 in PBS). They were then incubated for 2 minutes in 50 µg/ml proteinase K in PBST, and the digestion was stopped by incubating for 2 minutes in 2 mg/ml glycine in PBST. Embryos were washed two times for 5 minutes each in PBST, and re-fixed in PP for 20 minutes. Finally, embryos were washed 5 times for 5 minutes each in PBST. At this point embryos could be dehydrated in an ethanol series (30 %, 50 %, 70 % and 100% ethanol) and stored at -20 °C.

Hybridisation of embryos

To equilibrate into hybridisation solution (HYB: 50 % formamide, 5 x SSC, 100 µg/ml herring sperm DNA, 50 µg/ml heparin, 0.3 % Tween-20), 20 - 40 µl embryos were washed for 10 minutes in 50 % PBST, 50 % HYB, followed by 10 minutes in 100 % HYB. The embryos were then prehybridised for 1 - 3 hours at 45 °C in 200 µl HYB, which had been boiled and cooled on ice to denature the single stranded DNA.

Double stranded DNA probes were made by random priming using a digoxigenin labelling kit (Boehringer). 50 - 100 ng DNA fragment was labelled in a volume of 20 µl according to the manufacturers instructions, and the reaction was stopped by the addition of EDTA pH 8.0 to 10 mM. Typically, 4 µl of this probe was added to 25 µl HYB, boiled for 5 minutes and cooled on ice. This probe mixture was then added to the embryos (as much prehybridisation solution as possible was removed) and hybridisation was allowed to proceed overnight at 45 °C.

After hybridisation, the embryos were washed three times for 20 minutes each at 48 °C, in pre-warmed HYB. They were then washed for 20 minutes at 48 °C in 60 % HYB, 40 % PBST, and for 20 minutes at 48 °C in 20 % HYB, 80 % PBST. Finally, a 10 minute washing step at room temperature in PBST was carried out.

Detection of signal

Embryos were blocked for 2 hours at 4 °C in PBST/0.1 % BSA. They were then incubated for 2 hours at 4 °C in preabsorbed antibody conjugate (anti-digoxigenin antibody conjugated to alkaline phosphatase, Boehringer DIG random prime DNA labelling and detection kit), diluted 1/4000 in PBST/0.1 % BSA. Embryos were then washed four times for 15 minutes each in PBST, followed by 3 times for 5 minutes each in 1 ml alkaline phosphatase buffer (100 mM Tris.HCl pH 9.5, 100 mM NaCl, 50 mM MgCl₂, 0.1 % Tween20). 4.4 µl 75 mg/ml NBT and 3.3 µl 50 mg/ml BCIP were added to the last wash and the reaction was allowed to proceed for 3 - 16 hours at room temperature.

Staining was stopped by washing the embryos three times in PBST. They were then dehydrated in an ethanol series (30 %, 50 %, 70 % and 100 % ethanol) and mounted in Euparal.

Preabsorption of the anti-digoxigenin antibody conjugate

50 µl of wild-type embryos were dechorionated, fixed and devitellinised as described above. They were transferred in methanol to an Eppendorf tube, and rehydrated through a methanol series into PBS (70 % methanol/30 % PBS; 50 % methanol/50 % PBS; 30 % methanol/70 % PBS; 100 % PBS). Embryos were blocked for 1 hour at 4 °C in PBST/0.1 % BSA, before incubating for a further 1 hour at 4 °C in anti-digoxigenin antibody conjugate, which was diluted 1/400 in PBST/0.1 % BSA. Embryos were centrifuged for 1 minute at top speed in a microfuge, and the preabsorbed antibody solution was taken and stored for up to two weeks at 4 °C.

- CHAPTER 8 -

References

- Adamson, A. L. and Shearn, A. (1996). Molecular genetic analysis of *Drosophila ash2*, a member of the trithorax group required for imaginal disc pattern formation. *Genetics* 144, 621-633.
- Akam, M. (1987). The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* 101, 1-22.
- Akam, M. E. and Martinez-Arias, A. (1985). The distribution of *Ultrabithorax* transcripts in *Drosophila* embryos. *EMBO J.* 4, 1689-1700.
- Akasaka, T., Kanno, M., Balling, R., Miezo, M. A., Taniguchi, M. and Koseki, H. (1996). A role for *mel-18*, a Polycomb group related vertebrate gene, during the anteroposterior specification of the axial skeleton. *Development* 122, 1513-1522.
- Alkema, M. J., Bronk, M., Verhoeven, E., Otte, A., van't Veer, L. J., Berns, A. and van Lohuizen, M. (1997). Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian Polycomb complex. *Genes Dev.* 11, 226-240.
- Alkema, M. J., van der Lugt, N. M. T., Bobeldijk, R. C., Berns, A. and van Lohuizen, M. (1995). Transformation of axial skeleton due to overexpression of *bmi-1* in transgenic mice. *Nature* 374, 724-727.
- Aparicio, O. M., Billington, B. L. and Gottschling, D. E. (1991). Modifiers of position effect are shared between telomeric and silent mating type loci in *S. cerevisiae*. *Cell* 66, 1279-1287.
- Ashburner, M. (1989). *Drosophila: A laboratory manual* (Cold Spring Harbor Laboratory Press).
- Baker, N. E. (1987). Molecular cloning of sequences from *wingless*, a segment polarity gene in *Drosophila*: the spatial distribution of a transcript in embryos. *EMBO J.* 6, 1765-1773.
- Bannister, A. J. and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. *Nature* 384, 641-643.
- Becker, P. B. (1994). The establishment of active promoters in chromatin. *Bioessays* 16, 541-547.

References

- Becker, P. B. and Wu, C. (1992). Cell-free system for assembly of transcriptionally repressed chromatin from *Drosophila* embryos. *Mol. Cell. Biol.* 12, 2241-2249.
- Bender, W., Akam, M. E., Karch, F., Beachy, P. A., Peifer, M., Spierer, P., Lewis, E. B. and Hogness, D. S. (1983). Molecular genetics of the bithorax complex in *Drosophila melanogaster*. *Science* 221, 23-29.
- Bhat, K. M., Farkas, G., Karch, F., Gyurkovics, H., Gausz, J. and Schedl, P. (1996). The GAGA factor is required in the early *Drosophila* embryo not only for transcriptional regulation but also for nuclear division. *Development* 122, 1113-1124.
- Bienz, M. and Müller, J. (1995). Transcriptional silencing of homeotic genes in *Drosophila*. *Bioessays* 17, 775-784.
- Biggin, M. D. and Tjian, R. (1988). Transcription factors that activate the *Ultrabithorax* promoter in developmentally staged extracts. *Cell* 53, 699-711.
- Blair, S. S. (1995). Compartments and appendage development in *Drosophila*. *Bioessays* 17, 299-309.
- Bornemann, D., Miller, E. and Simon, J. (1996). The *Drosophila* Polycomb group gene *Sex comb on midleg (Scm)* encodes a zinc finger protein with similarity to polyhomeotic protein. *Development* 122, 1621-1630.
- Breen, T. R., Chinwalla, V. and Harte, P. J. (1995). *Trithorax* is required to maintain *engrailed* expression in a subset of *engrailed*-expressing cells. *Mech. Dev.* 52, 89-98.
- Breen, T. R. and Duncan, I. M. (1986). Maternal expression of genes that regulate the bithorax complex of *Drosophila melanogaster*. *Dev. Biol.* 118, 442-456.
- Breen, T. R. and Harte, P. J. (1993). *trithorax* regulates multiple homeotic genes in the bithorax and Antennapedia complexes and exerts different tissue-specific, parasegment-specific and promoter-specific effects on each. *Development* 117, 119-134.
- Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y. and Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: a homologue to yeast Gcn5p linking histone acetylation to gene activation. *Cell* 84, 843-851.
- Brunk, B. P., Martin, E. C. and Adler, P. N. (1991). *Drosophila* genes *Posterior sex combs* and *Suppressor two of zeste* encode proteins with homology to the murine *bmi-1* oncogene. *Nature* 353, 351-353.

References

- Buchenau, P., Hodgson, J., Strutt, H. and Arndt-Jovin, D. J. (1997). The distribution of Polycomb-group proteins during cell division and development in *Drosophila* embryos: impact on models for silencing. Manuscript submitted.
- Bunker, C. A. and Kingston, R. E. (1994). Transcriptional repression by *Drosophila* and mammalian Polycomb group proteins in transfected mammalian cells. *Mol. Cell. Biol.* 14, 1721-1732.
- Busturia, A. and Bienz, M. (1993). Silencers in *Abdominal-B*, a homeotic *Drosophila* gene. *EMBO J.* 12, 1415-1425.
- Busturia, A. and Morata, G. (1988). Ectopic expression of homeotic genes caused by the elimination of the *Polycomb* gene in *Drosophila* imaginal epidermis. *Development* 104, 713-720.
- Cai, H. N. and Levine, M. (1997). The gypsy insulator can function as a promoter-specific silencer in the *Drosophila* embryo. *EMBO J.* 16, 1732-1741.
- Cairns, B. R., Kim, Y.-J., Sayre, M. H., Laurent, B. C. and Kornberg, R. D. (1994). A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5 and SNF6 gene products isolated from yeast. *Proc. Natl. Acad. Sci. USA* 91, 1950-1954.
- Cairns, B. R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B. and Kornberg, R. D. (1996). RSC, an essential, abundant chromatin-remodeling complex. *Cell* 87, 1249-1260.
- Campbell, R. B., Sinclair, D. A. R., Couling, M. and Brock, H. W. (1995). Genetic interactions and dosage effects of Polycomb group genes of *Drosophila*. *Mol. Gen. Genet.* 246, 291-300.
- Carrington, E. A. and Jones, R. S. (1996). The *Drosophila* Enhancer of *zeste* gene encodes a chromosomal protein: examination of wild-type and mutant protein distribution. *Development* 122, 4073-4083.
- Casares, F. and Sanchez-Herrero, E. (1995). Regulation of the *infraabdominal* regions of the bithorax complex of *Drosophila* by gap genes. *Development* 121, 1855-1866.
- Celniker, S. E., Sharma, S., Keelan, D. J. and Lewis, E. B. (1990). The molecular genetics of the bithorax complex of *Drosophila*: cis-regulation in the *Abdominal-B* domain. *EMBO J.* 9, 4277-4286.

References

- Chan, C.-S., Rastelli, L. and Pirrotta, V. (1994). A *Polycomb* response element in the *Ubx* gene that determines an epigenetically inherited state of repression. *EMBO J.* 13, 2553-2564.
- Chang, Y.-L., King, B. O., O'Connor, M. O., Mazo, A. and Huang, D.-H. (1995). Functional reconstruction of *trans* regulation of the *Ultrabithorax* promoter by the products of two antagonistic genes, *trithorax* and *Polycomb*. *Mol. Cell. Biol.* 15, 6601-6612.
- Cheng, N. N., Sinclair, D. A. R., Campbell, R. B. and Brock, H. W. (1994). Interactions of *polyhomeotic* with Polycomb group genes of *Drosophila melanogaster*. *Genetics* 138, 1151-1162.
- Chiang, A., O'Connor, M. B., Paro, R., Simon, J. and Bender, W. (1995). Discrete Polycomb-binding sites in each parasegmental domain of the bithorax complex. *Development* 121, 1681-1689.
- Chien, C.-t., Buck, S., Sternglanz, R. and Shore, D. (1993). Targeting of SIR1 protein establishes transcriptional silencing at *HM* loci and telomeres in yeast. *Cell* 75, 531-541.
- Chinwalla, V., Jane, E. P. and Harte, P. J. (1995). The *Drosophila trithorax* protein binds to specific chromosomal sites and is co-localized with *Polycomb* at many sites. *EMBO J.* 14, 2056-2065.
- Chouinard, S. and Kaufman, T. C. (1991). Control of expression of the homeotic *labial (lab)* locus of *Drosophila melanogaster*: evidence for both positive and negative autogenous regulation. *Development* 113, 1267-1280.
- Christen, B. and Bienz, M. (1994). Imaginal disc silencers from *Ultrabithorax*: evidence for Polycomb response elements. *Mech. Dev.* 48, 255-266.
- Church, G. M. and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* 81, 1991-1995.
- Coleman, K. G., Poole, S. J., Weir, M. P., Soeller, W. C. and Kornberg, T. (1987). The *invected* gene of *Drosophila*: sequence analysis and expression studies reveal a close kinship to the *engrailed* gene. *Genes Dev.* 1, 19-28.
- Corces, V. G. and Geyer, P. K. (1991). Interactions of retrotransposons with the host genome: the case of the *gypsy* element of *Drosophila*. *TIG* 7, 86-90.
- Coré, M., Bel, S., Gaunt, S. J., Aurrand-Lions, M., Pearce, J., Fisher, A. and Djabali, M. (1997). Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. *Development* 124, 721-729.

References

- Cote, J., Quinn, J., Workman, J. L. and Peterson, C. L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science* 265, 53-60.
- Craig, E. A. (1980). Four *Drosophila* heat shock genes at 67B: characterisation of recombinant plasmids. *Nucl. Acids Res.* 8, 4441-4457.
- Croston, G. E., Kerrigan, L. A., Lira, L. M., Marshak, D. R. and Kadonaga, J. T. (1991). Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. *Science* 251, 643-649.
- Csink, A. K. and Henikoff, S. (1996). Genetic modification of heterochromatic association and nuclear organisation in *Drosophila*. *Nature* 381, 529-531.
- Cumberledge, S., Zaratzian, A. and Sakonju, S. (1990). Characterisation of two RNAs transcribed from the cis-regulatory region of the *abd-A* domain within the *Drosophila* bithorax complex. *Proc. Natl. Acad. Sci. USA* 87, 3259-3263.
- De Robertis, F., Kadosh, D., Henchoz, S., Pauli, D., Reuter, G., Struhl, K. and Spierer, P. (1996). The histone deacetylase RPD3 counteracts genomic silencing in *Drosophila* and yeast. *Nature* 384, 589-591.
- DeCamillis, M., Cheng, N., Pierre, D. and Brock, H. W. (1992). The *polyhomeotic* gene of *Drosophila* encodes a chromatin protein that shares polytene chromosome-binding sites with *Polycomb*. *Genes Dev.* 6, 223-232.
- Denell, R. E. and Frederick, R. D. (1983). Homeosis in *Drosophila*: a description of the Polycomb lethal syndrome. *Dev. Biol.* 97, 34-47.
- Dernburg, A. F., Broman, K. W., Fung, J. C., Marshall, W. F., Philips, J., Agard, D. A. and Sedat, J. W. (1996). Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell* 85, 745-759.
- DiNardo, S., Sher, E., Heemskerk-Jongens, J., Kassis, J. A. and O'Farrell, P. H. (1988). Two-tiered regulation of spatially patterned *engrailed* gene expression during *Drosophila* embryogenesis. *Nature* 332, 604-609.
- Dingwall, A. K., Beek, S. J., McCallum, C. M., Tamkun, J. W., Kalpana, G. V., Goff, S. P. and Scott, M. P. (1995). The *Drosophila* snr1 and brm proteins are related to the yeast SWI/SNF proteins and are components of a large protein complex. *Mol. Biol. Cell* 6, 777-791.

References

- Docquier, F., Saget, O., Forquignon, F., Randsholt, N. B. and Santamaria, P. (1996). *multi sex combs*, a Polycomb group gene of *Drosophila melanogaster*, is required for proliferation of the germline. *Roux's Arch. Dev. Biol.* 205, 203-214.
- Dorer, D. R. and Henikoff, S. (1994). Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. *Cell* 77, 993-1002.
- Drees, B., Ali, Z., Soeller, W. C., Coleman, K. C., Poole, S. J. and Kornberg, T. (1987). The transcription unit of the *Drosophila engrailed* locus: an unusually small portion of a 70 000 bp gene. *EMBO J.* 6, 2803-2809.
- Driever, W. and Nüsslein-Volhard, C. (1988). The *bicoid* protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell* 54, 95-104.
- Driever, W. and Nüsslein-Volhard, C. (1989). The *bicoid* protein is a positive regulator of *hunchback* transcription in the early *Drosophila* embryo. *Nature* 337, 138-143.
- Duncan, I. (1986). Control of bithorax complex functions by the segmentation gene *fushi tarazu* of *D. melanogaster*. *Cell* 47, 297-309.
- Dura, J.-M. and Ingham, P. (1988). Tissue- and stage-specific control of homeotic and segmentation gene expression in *Drosophila* embryos by the *polyhomeotic* gene. *Development* 103, 733-741.
- Dura, J.-M., Randsholt, N. B., Deatrick, J., Erk, I., Santamaria, P., Freeman, J. D., Freeman, S. J., Weddell, D. and Brock, H. W. (1987). A complex genetic locus, *polyhomeotic*, is required for segmental specification and epidermal development in *Drosophila melanogaster*. *Cell* 51, 829-839.
- Eissenberg, J. C. and Elgin, S. C. R. (1991). Boundary functions in the control of gene expression. *TIG* 7, 335-340.
- Eissenberg, J. C., Elgin, S. C. R. and Paro, R. (1995). Epigenetic regulation in *Drosophila*: a conspiracy of silence. In *Chromatin structure and gene expression*, Elgin, S. C. R., ed. (Oxford University Press), pp. 147-171.
- Eissenberg, J. C., James, T. C., Foster-Hartnett, D. M., Hartnett, T., Ngan, V. and Elgin, S. C. R. (1990). Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 87, 9923-9927.

References

- Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H. and Karch, F. (1994). The *Trithorax-like* gene encodes the *Drosophila* GAGA factor. *Nature* 371, 806-808.
- Fauvarque, M.-O. and Dura, J.-M. (1993). *polyhomeotic* regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in *Drosophila*. *Genes Dev.* 7, 1508-1520.
- Fauvarque, M.-O., Zuber, V. and Dura, J.-M. (1995). Regulation of *polyhomeotic* transcription may involve local changes in chromatin activity in *Drosophila*. *Mech. Dev.* 52, 343-355.
- Fjose, A., McGinnis, W. J. and Gehring, W. J. (1985). Isolation of a homeobox-containing gene from the *engrailed* region of *Drosophila* and the spatial distribution of its transcripts. *Nature* 313, 284-289.
- Franke, A., DeCamillis, M., Zink, D., Cheng, N., Brock, H. W. and Paro, R. (1992). *Polycomb* and *polyhomeotic* are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*. *EMBO J.* 11, 2941-2950.
- Franke, A., Messmer, S. and Paro, R. (1995). Mapping functional domains of the Polycomb protein of *Drosophila melanogaster*. *Chrom. Res.* 3, 351-360.
- Galloni, M., Gyurkovics, H., Schedl, P. and Karch, F. (1993). The bluetail transposon: evidence for independent *cis*-regulatory domains and domain boundaries in the bithorax complex. *EMBO J.* 12, 1087-1097.
- Garcia-Bellido, A. (1975). Genetic control of wing disc development in *Drosophila*. In *Cell patterning: Ciba Foundation Symposium* 29, 161-178.
- Garcia-Bellido, A. (1977). Homeotic and atavic mutations in insects. *Am. Zool.* 17, 613-629.
- Gaul, U. and Jäckle, H. (1987). Pole region-dependent repression of the *Drosophila* gap gene *Krüppel* by maternal gene products. *Cell* 51, 549-555.
- Gerasimova, T. I., Gdula, D. A., Gerasimov, D. V., Simonova, O. and Corces, V. G. (1995). A *Drosophila* protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell* 82, 587-597.
- Geyer, P. K. (1997). The role of insulator elements in defining domains of gene expression. *Curr. Opin. Genet. Dev.* 7, 242-248.

References

- Gilmour, D. S. and Lis, J. T. (1986). RNA polymerase II interacts with the promoter region of the non-induced *hsp70* gene in *Drosophila melanogaster* cells. *Mol. Cell. Biol.* 6, 3984-3989.
- Gilmour, D. S., Thomas, G. H. and Elgin, S. C. R. (1989). *Drosophila* nuclear proteins bind to regions of alternating C and T residues in gene promoters. *Science* 245, 1487-1490.
- Gindhart, J. G. and Kaufmann, T. C. (1995). Identification of *Polycomb* and *trithorax* group responsive elements in the regulatory region of the *Drosophila* homeotic gene *Sex combs reduced*. *Genetics* 139, 797-814.
- Girton, J. R. and Jeon, S. H. (1994). Novel embryonic and adult homeotic phenotypes are produced by *pleiohomeotic* mutations in *Drosophila*. *Dev. Biol.* 161, 393-407.
- Glicksman, M. A. and Brower, D. L. (1990). Persistent ectopic expression of *Drosophila* homeotic genes resulting from maternal deficiency of the *extra sex combs* gene product. *Dev. Biol.* 142, 422-431.
- Goto, T., Macdonald, P. and Maniatis, T. (1989). Early and late periodic patterns of *even skipped* expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* 57, 413-422.
- Gottschling, D. E., Aparicio, O. M., Billington, B. L. and Zakian, V. A. (1990). Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* 63, 751-762.
- Gould, A. P., Brookman, J. J., Strutt, D. I. and White, R. A. H. (1990). Targets of homeotic gene control in *Drosophila*. *Nature* 348, 308-312.
- Gould, A. P. and White, R. A. H. (1992). *Connectin*, a target of homeotic gene control in *Drosophila*. *Development* 116, 1163-1174.
- Granok, H., Leibovitch, B. A., Shaffer, C. D. and Elgin, S. C. R. (1995). Ga-ga over GAGA factor. *Curr. Biol.* 5, 238-241.
- Gunster, M. J., Satijn, D. P. E., Hamer, K. M., den Blaauwen, J. L., de Bruijn, D., Alkema, M. J., van Lohuizen, M., van Driel, R. and Otte, A. P. (1997). Identification and characterisation of interactions between the vertebrate Polycomb-group protein BMI1 and human homologues of Polyhomeotic. *Mol. Cell. Biol.* 17, 2326-2335.
- Gustavson, E., Goldsborough, A. S., Ali, Z. and Kornberg, T. B. (1996). The *Drosophila engrailed* and *invected* genes: partners in regulation, expression and function. *Genetics* 142, 893-906.

References

- Gutjahr, T., Frei, E., Spicer, C., Baumgartner, S., White, R. A. H. and Noll, M. (1995). The Polycomb-group gene, *extra sex combs*, encodes a nuclear member of the WD-40 repeat family. *EMBO J.* 14, 4296-4306.
- Gyurkovics, H., Gausz, J., Kummer, J. and Karch, F. (1990). A new homeotic mutation in the *Drosophila* bithorax complex removes a boundary separating two domains of regulation. *EMBO J.* 9, 2579-2585.
- Hafen, E., Levine, M. and Gehring, W. J. (1984). Regulation of *Antennapedia* transcript distribution by the bithorax complex in *Drosophila*. *Nature* 307, 287-289.
- Hagstrom, K., Muller, M. and Schedl, P. (1996). *Fab-7* functions as a chromatin domain boundary to ensure proper segment specification by the *Drosophila* bithorax complex. *Genes Dev.* 10, 3202-3215.
- Hama, C., Ali, Z. and Kornberg, T. B. (1990). Region-specific recombination and expression are directed by portions of the *Drosophila engrailed* promoter. *Genes Dev.* 4, 1079-1093.
- Harding, K., Hoey, T., Warrior, R. and Levine, M. (1989). Autoregulatory and gap response elements of the *even-skipped* promoter of *Drosophila*. *EMBO J.* 8, 1205-1212.
- Harding, K. and Levine, M. (1988). Gap genes define the limits of *Antennapedia* and *Bithorax* gene expression during early development in *Drosophila*. *EMBO J.* 7, 205-214.
- Harding, K., Wedeen, C., McGinnis, W. and Levine, M. (1985). Spatially regulated expression of homeotic genes in *Drosophila*. *Science* 229, 1236-1242.
- Harlow, E. and Lane, D. (1988). *Antibodies: a laboratory manual* (Cold Spring Harbor Laboratory Press).
- Hecht, A., LaRoche, T., Strahl-Bolsinger, S., Gasser, S. M. and Grunstein, M. (1995). Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. *Cell* 80, 583-592.
- Hecht, A., Strahl-Bolsinger, S. and Grunstein, M. (1996). Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. *Nature* 383, 92-96.
- Heemskerk, J., DiNardo, S., Kostriken, R. and O'Farrell, P. H. (1991). Multiple modes of *engrailed* regulation in the progression towards cell fate determination. *Nature* 352, 404-410.
- Henikoff, S. (1994). A reconsideration of the mechanism of position effect. *Genetics* 138, 1-5.

References

- Henikoff, S. (1996). Dosage-dependent modification of position-effect variegation in *Drosophila*. *Bioessays* 18, 401-409.
- Hidalgo, A. (1996). The roles of *engrailed*. *TIG* 12, 1-4.
- Hirschhorn, J. N., Brown, S. A., Clark, C. D. and Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes Dev.* 6, 2288-2298.
- Hobert, O., Sures, I., Ciossek, T., Fuchs, M. and Ullrich, A. (1996). Isolation and developmental expression analysis of *Enx-1*, a novel mouse Polycomb group gene. *Mech. Dev.* 55, 171-184.
- Howard, K. and Ingham, P. (1986). Regulatory interactions between the segmentation genes *fushi tarazu*, *hairy*, and *engrailed* in the *Drosophila* blastoderm. *Cell* 44, 949-957.
- Howard, K., Ingham, P. and Rushlow, C. (1988). Region-specific alleles of the *Drosophila* segmentation gene *hairy*. *Genes Dev.* 2, 1037-1046.
- Imbalzano, A. N., Kwon, H., Green, M. R. and Kingston, R. E. (1994). Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature* 370, 481-485.
- Ingham, P. W. (1983). Differential expression of bithorax complex genes in absence of the *extra sex combs* and *trithorax* genes. *Nature* 306, 591-593.
- Ingham, P. W. (1984). A gene that regulates the bithorax complex differentially in larval and adult cells of *Drosophila*. *Cell* 37, 815-823.
- Ingham, P. W. (1985). Genetic control of the spatial pattern of selector gene expression in *Drosophila*. *Cold Spring Harbor Symp. Quant. Biol.* 50, 201-208.
- Ingham, P. W. (1988). The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature* 335, 25-34.
- Ingham, P. W., Ish-Horowicz, D. and Howard, K. R. (1986). Correlative changes in homeotic and segmentation gene expression in *Krüppel* mutant embryos of *Drosophila*. *EMBO J.* 5, 1659-1665.
- Ingham, P. W. and Martinez-Arias, A. (1986). The correct activation of Antennapedia and bithorax complex genes requires the *fushi tarazu* gene. *Nature* 324, 592-597.

References

- Ingham, P. W. and Whittle, R. (1980). *Trithorax*: a new homeotic mutation of *Drosophila melanogaster* causing transformations of abdominal and thoracic segments. I. Putative role during embryogenesis. *Mol. Gen. Genet.* 179, 607-614.
- Irish, V. F., Martinez-Arias, A. and Akam, M. (1989). Spatial regulation of the *Antennapedia* and *Ultrabithorax* homeotic genes during *Drosophila* early development. *EMBO J.* 8, 1527-1537.
- Irvine, K. D., Helfand, S. L. and Hogness, D. S. (1991). The large upstream control region of the *Drosophila* homeotic gene *Ultrabithorax*. *Development* 111, 407-424.
- Jack, T., Regulski, M. and McGinnis, W. (1988). Pair-rule segmentation genes regulate the expression of the homeotic selector gene *Deformed*. *Genes Dev.* 2, 635-651.
- Jäckle, H., Tautz, D., Schuh, R., Seifert, E. and Lehmann, R. (1986). Cross-regulatory interactions among the gap genes of *Drosophila*. *Nature* 324, 668-670.
- James, T. C. and Elgin, S. C. R. (1986). Identification of a nonhistone chromosomal protein associated with heterochromatin in *Drosophila melanogaster* and its gene. *Mol. Cell. Biol.* 6, 3862-3872.
- Jones, R. S. and McGinnis, W. (1993). The regulation of *empty spiracles* by *Abdominal-B* mediates an abdominal segment identity function. *Genes Dev.* 7, 229-240.
- Jones, R. S. and Gelbart, W. M. (1993). The *Drosophila* Polycomb-group gene *Enhancer of zeste* contains a region with sequence similarity to *trithorax*. *Mol. Cell. Biol.* 13, 6357-6366.
- Jürgens, G. (1985). A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*. *Nature* 316, 153-155.
- Kanno, M., Hasegawa, M., Ishida, A., Isono, K. and Taniguchi, M. (1995). *mel-18*, a Polycomb group-related mammalian gene, encodes a transcriptional negative regulator with tumor suppressive activity. *EMBO J.* 14, 5672-5678.
- Karch, F., Galloni, M., Sipos, L., Gausz, J., Gyurkovics, H. and Schedl, P. (1994). *Mcp* and *Fab-7*: molecular analysis of putative boundaries of *cis*-regulatory domains in the bithorax complex of *Drosophila melanogaster*. *Nucl. Acids Res.* 22, 3138-3146.
- Karch, F., Weiffenbach, B., Peifer, M., Bender, W., Duncan, I., Celniker, S., Crosby, M. and Lewis, E. B. (1985). The abdominal region of the bithorax complex. *Cell* 43, 81-96.

References

- Kassis, J. A. (1990). Spatial and temporal control elements of the *Drosophila engrailed* gene. *Genes Dev.* 4, 433-443.
- Kassis, J. A. (1994). Unusual properties of regulatory DNA from the *Drosophila engrailed* gene: three "pairing-sensitive" sites within a 1.6-kb region. *Genetics* 136, 1025-1038.
- Kassis, J. A., VanSickle, E. P. and Sensabaugh, S. M. (1991). A fragment of *engrailed* regulatory DNA can mediate transvection of the *white* gene in *Drosophila*. *Genetics* 128, 751-761.
- Kaufman, T., Seeger, M. A. and Olsen, G. (1990). The molecular and genetic organisation of the Antennapedia gene complex of *Drosophila melanogaster*. In *Advances in Genetics* 27, Wright, T. R. F. and Scandalios, J. G., eds. (Academic press), pp. 309-362.
- Kayne, P. S., Kim, U.-J., Han, M., Mullen, J. R., Yoshizaki, F. and Grunstein, M. (1988). Extremely conserved histone H4 N-terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. *Cell* 55, 27-39.
- Kellum, R. and Alberts, B. M. (1995). Heterochromatin protein 1 is required for correct chromosome segregation in *Drosophila* embryos. *J. Cell Sci.* 108, 1419-1431.
- Kellum, R. and Schedl, P. (1991). A position effect assay for boundaries of higher order chromosomal domains. *Cell* 64, 941-950.
- Kellum, R. and Schedl, P. (1992). A group of scs elements function as domain boundaries in an enhancer blocking assay. *Mol. Cell. Biol.* 12, 2424-2431.
- Kennison, J. A. and Russell, M. A. (1987). Dosage-dependent modifiers of homeotic mutations in *Drosophila melanogaster*. *Genetics* 116, 75-86.
- Kennison, J. A. and Tamkun, J. W. (1988). Dosage-dependent modifiers of *Polycomb* and *Antennapedia* mutations in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 85, 8136-8140.
- Khavari, P. A., Peterson, C. L., Tamkun, J. W., Mendel, D. B. and Crabtree, G. R. (1993). BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature* 366, 170-174.
- Kingston, R. E., Bunker, C. A. and Imbalzano, A. N. (1996). Repression and activation by multiprotein complexes that alter chromatin structure. *Genes Dev.* 10, 905-920.

References

- Kornberg, T., Siden, I., O'Farrell, P. and Simon, M. (1985). The *engrailed* locus of *Drosophila*: *in situ* localization of the transcripts reveals compartment specific expression. *Cell* 40, 45-53.
- Kruger, W. and Herskowitz, I. (1991). A negative regulator of HO transcription, SIN1 (SPT2), is a nonspecific DNA binding protein related to HMG1. *Mol. Cell. Biol.* 11, 4135-4146.
- Kruger, W., Peterson, C. L., Sil, A., Coburn, C., Arents, G., Moudrianakis, E. N. and Herskowitz, I. (1995). Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. *Genes Dev.* 9, 2770-2779.
- Krumlauf, R. (1994). *Hox* genes in vertebrate development. *Cell* 78, 191-201.
- Kuner, J. M., Nakanishi, M., Ali, Z., Drees, B., Gustavson, E., Theis, J., Kauver, L., Kornberg, T. and O'Farrell, P. H. (1985). Molecular cloning of *engrailed*: a gene involved in the development of pattern in *Drosophila melanogaster*. *Cell* 42, 309-316.
- Kuroiwa, A., Kloter, U., Baumgartner, P. and Gehring, W. J. (1985). Cloning of the homeotic *Sex combs reduced* gene in *Drosophila* and *in situ* localisation of its transcripts. *EMBO J.* 4, 3757-3764.
- Kuzin, B., Tillib, S., Sedkov, Y., Mizrokhi, L. and Mazo, A. (1994). The *Drosophila trithorax* gene encodes a chromosomal protein and directly regulates the region-specific homeotic gene *forkhead*. *Genes Dev.* 8, 2478-2490.
- Kuziora, M. A. and McGinnis, W. (1988). Autoregulation of a *Drosophila* homeotic selector gene. *Cell* 55, 477-485.
- Kwon, H., Imbalzano, A. N., Khavari, P. A., Kingston, R. E. and Green, M. R. (1994). Nucleosome disruption and enhancement of activator binding by a human SWI/SNF complex. *Nature* 370, 477-481.
- LaJeunesse, D. and Shearn, A. (1995). Trans-regulation of thoracic homeotic selector genes of the Antennapedia and bithorax complexes by the trithorax group genes: *absent*, *small*, and *homeotic discs 1* and *2*. *Mech. Dev.* 53, 123-139.
- LaJeunesse, D. and Shearn, A. (1996). *E(z)*: a Polycomb group gene or a trithorax group gene? *Development* 122, 2189-2197.
- Landecker, H. L., Sinclair, D. A. R. and Brock, H. W. (1994). Screen for enhancers of *Polycomb* and *Polycomblike* in *Drosophila melanogaster*. *Dev. Genet.* 15, 425-434.

References

- Laurenson, P. and Rine, J. (1992). Silencers, silencing and heritable transcriptional states. *Microbiol. Rev.* 56, 543-560.
- Laurent, B. C., Treich, I. and Carlson, M. (1993). The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. *Genes Dev.* 7, 583-591.
- Lawrence, P. A. and Morata, G. (1976). Compartments in the wing of *Drosophila*: a study of the *engrailed* gene. *Dev. Biol.* 50, 321-337.
- Lawrence, P. A. and Morata, G. (1994). Homeobox genes: their function in *Drosophila* segmentation and pattern formation. *Cell* 78, 181-189.
- Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* 276, 565-570.
- Lindsley, D. L. and Zimm, G. G. (1992). The genome of *Drosophila melanogaster* (Academic Press).
- Locke, J., Kotarski, M. A. and Tartof, K. D. (1988). Dosage-dependent modifiers of position effect variegation in *Drosophila* and a mass action model that explains their effect. *Genetics* 120, 181-198.
- Lonie, A., D'Andrea, R., Paro, R. and Saint, R. (1994). Molecular characterisation of the *Polycomb-like* gene of *Drosophila melanogaster*, a *trans*-acting negative regulator of homeotic gene expression. *Development* 120, 2629-2636.
- Lu, Q., Wallrath, L. L., Granok, H. and Elgin, S. C. R. (1993a). (CT)_n.(GA)_n repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the *Drosophila hsp26* gene. *Mol. Cell. Biol.* 13, 2802-2814.
- Lu, X., Perkins, L. A. and Perrimon, N. (1993b). The *torso* pathway in *Drosophila*: a model system to study receptor tyrosine kinase signal transduction. *Development (Suppl.)*, 47-56.
- Lüdecke, H.-J., Senger, G., Claussen, U. and Horsthemke, B. (1989). Cloning defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification. *Nature* 338, 348-350.
- Maillet, L., Boscheron, C., Gotta, M., Marcand, S., Gilson, E. and Gasser, S. M. (1996). Evidence for silencing compartments within the yeast nucleus: a role for telomere proximity and Sir protein concentration in silencer-mediated repression. *Genes Dev.* 10, 1796-1811.

References

- Mann, R. S. (1994). *engrailed*-mediated repression of *Ultrabithorax* is necessary for the parasegment 6 identity in *Drosophila*. *Development* 120, 3205-3212.
- Marcand, S., Buck, S. W., Moretti, P., Gilson, E. and Shore, D. (1996a). Silencing of genes at nontelomeric sites in yeast is controlled by sequestration of silencing factors at telomeres by Rap1 protein. *Genes Dev.* 10, 1297-1309.
- Marcand, S., Gasser, S. M. and Gilson, E. (1996b). Chromatin: a sticky silence. *Curr. Biol.* 6, 1222-1225.
- Martin, C. H., Mayeda, C. A., Davis, C. A., Ericsson, C. L., Knafels, J. D., Mathog, D. R., Celniker, S. E., Lewis, E. B. and Palazzolo, M. J. (1995). Complete sequence of the bithorax complex of *Drosophila*. *Proc. Natl. Acad. Sci. USA* 92, 8398-8402.
- Martin, E. C. and Adler, P. N. (1993). The *Polycomb* group gene *Posterior sex combs* encodes a chromosomal protein. *Development* 117, 641-655.
- Martinez-Arias, A., Baker, N. E. and Ingham, P. W. (1988). Role of segment polarity genes in the definition and maintenance of cell states in the *Drosophila* embryo. *Development* 103, 157-170.
- Martinez-Arias, A. and White, R. A. H. (1988). *Ultrabithorax* and *engrailed* expression in *Drosophila* embryos mutant for segmentation genes of the pair-rule class. *Development* 102, 325-338.
- Mazo, A. M., Huang, D.-H., Mozer, B. A. and Dawid, I. B. (1990). The *trithorax* gene, a trans-acting regulator of the bithorax complex in *Drosophila*, encodes a protein with zinc-binding domains. *Proc. Natl. Acad. Sci. USA* 87, 2112-2116.
- McCall, K. and Bender, W. (1996). Probes for chromatin accessibility in the *Drosophila* bithorax complex respond differently to *Polycomb*-mediated repression. *EMBO J.* 15, 569-580.
- McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* 68, 283-302.
- McGinnis, W., Levine, M. S., Hafen, E., Kuriowa, A. and Gehring, W. J. (1984). A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature* 308, 428-433.
- McKeon, J. and Brock, H. W. (1991). Interactions of the *Polycomb* group of genes with homeotic loci of *Drosophila*. *Roux's Arch. Dev. Biol.* 199, 387-396.

References

- McKeon, J., Slade, E., Sinclair, D. A. R., Cheng, N., Couling, M. and Brock, H. W. (1994). Mutations in some *Polycomb* group genes of *Drosophila* interfere with regulation of segmentation genes. *Mol. Gen. Genet.* 244, 474-483.
- Merrill, V. K. L., Diederich, R. J., Turner, F. R. and Kaufman, T. C. (1989). A genetic and developmental analysis of mutations in *labial*, a gene necessary for proper head formation in *Drosophila melanogaster*. *Dev. Biol.* 135, 376-391.
- Messmer, S., Franke, A. and Paro, R. (1992). Analysis of the functional role of the *Polycomb* chromodomain in *Drosophila melanogaster*. *Genes Dev.* 6, 1241-1254.
- Mihaly, J., Hogga, I., Gausz, J., Gyurkovics, H. and Karch, F. (1997). In situ dissection of the *Fab-7* region of the bithorax complex into a chromatin domain boundary and a Polycomb response element. *Development* 124, 1809-1820.
- Mizzen, C. A., Yang, X.-J., Kokubo, T., Brownell, J. E., Bannister, A. J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S. L., Kouzarides, T., Nakatani, Y. and Allis, C. D. (1996). The TAF_{II}250 subunit of TFIID has histone acetyltransferase activity. *Cell* 87, 1261-1270.
- Moazed, D. and O'Farrell, P. H. (1992). Maintenance of the *engrailed* expression pattern by *Polycomb* group genes in *Drosophila*. *Development* 116, 805-810.
- Morata, G. and Lawrence, P. A. (1975). Control of compartment development by the *engrailed* gene in *Drosophila*. *Nature* 255, 614-617.
- Moretti, P., Freeman, K., Coodly, L. and Shore, D. (1994). Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. *Genes Dev.* 8, 2257-2269.
- Morisato, D. and Anderson, K. V. (1995). Signaling pathways that establish the dorsal-ventral pattern of the *Drosophila* embryo. *Annu. Rev. Genetics* 29, 371-399.
- Muchardt, C. and Yaniv, M. (1993). A human homologue of *Saccharomyces cerevisiae* SNF2/SWI2 and *Drosophila* *brm* genes potentiates transcriptional activation by the glucocorticoid receptor. *EMBO J.* 12, 4279-4290.
- Müller, J. (1995). Transcriptional silencing by the Polycomb protein in *Drosophila* embryos. *EMBO J.* 14, 1209-1220.
- Müller, J. and Bienz, M. (1991). Long range repression conferring boundaries of *Ultrabithorax* expression in the *Drosophila* embryo. *EMBO J.* 10, 3147-3155.

References

- Müller, J., Gaunt, S. and Lawrence, P. A. (1995). Function of the Polycomb protein is conserved in mice and flies. *Development* 121, 2847-2852.
- Nomura, M., Takiara, Y. and Shimada, K. (1994). Isolation and characterisation of retinoic acid-inducible cDNA clones in F9 cells: one of the early inducible clones encodes a novel protein sharing several highly homologous regions with a *Drosophila polyhomeotic* protein. *Differentiation* 57, 39-50.
- Nüsslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287, 795-801.
- O'Brien, T., Wilkins, R. C., Giardina, C. and Lis, J. T. (1995). Distribution of GAGA protein on *Drosophila* genes *in vivo*. *Genes Dev.* 9, 1098-1110.
- Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* 87, 953-959.
- Omichinski, J. G., Pedone, P. V., Felsenfeld, G., Gronenborn, A. M. and Clore, G. M. (1997). The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nature Struct. Biol.* 4, 122-132.
- Orlando, V. and Paro, R. (1993). Mapping Polycomb-repressed domains in the bithorax complex using *in vivo* formaldehyde cross-linked chromatin. *Cell* 75, 1187-1198.
- Orlando, V., Strutt, H. and Paro, R. (1997). Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. *METHODS: A Companion to Methods in Enzymol.* 11, 205-214.
- Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L. and Gasser, S. M. (1993). SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. *Cell* 75, 543-555.
- Pankratz, M. J., Hoch, M., Seifert, E. and Jäckle, H. (1989). *Krüppel* requirement for *knirps* enhancement reflects overlapping gap gene activities in the *Drosophila* embryo. *Nature* 341, 337-340.
- Pankratz, M. J., Seifert, E., Gerwin, N., Billi, B., Nauber, U. and Jäckle, H. (1990). Gradients of *Krüppel* and *knirps* gene products direct pair-rule gene stripe patterning in the posterior region of the *Drosophila* embryo. *Cell* 61, 309-317.
- Paro, R. (1990). Imprinting a determined state into the chromatin of *Drosophila*. *TIG* 6, 416-421.

References

- Paro, R. (1993). Mechanisms of heritable gene repression during development of *Drosophila*. *Curr. Opin. Cell Biol.* 5, 999-1005.
- Paro, R. (1995). Propagating memory of transcriptional states. *TIG* 11, 295-297.
- Paro, R. and Harte, P. J. (1996). The role of Polycomb group and trithorax group chromatin complexes in the maintenance of determined cell states. In *Epigenetic mechanisms of gene regulation*, Russo, V. E. A., Martienssen, R. A. and Riggs, A. D., eds. (Cold Spring Harbor Laboratory Press), pp. 507-528.
- Paro, R. and Hogness, D. S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proc. Natl. Acad. Sci. USA* 88, 263-267.
- Paro, R. and Zink, B. (1992). The *Polycomb* gene is differentially regulated during oogenesis and embryogenesis of *Drosophila melanogaster*. *Mech. Dev.* 40, 37-46.
- Pearce, J. J. H., Singh, P. B. and Gaunt, S. J. (1992). The mouse has a Polycomb-like chromobox gene. *Development* 114, 921-929.
- Peifer, M., Karch, F. and Bender, W. (1987). The bithorax complex: control of segmental identity. *Genes Dev.* 1, 891-898.
- Pelegri, F. and Lehmann, R. (1994). A role of Polycomb group genes in the regulation of gap gene expression in *Drosophila*. *Genetics* 136, 1341-1353.
- Perrimon, N. (1995). Hedgehog and beyond. *Cell* 80, 517-520.
- Peterson, C. L., Dingwall, A. and Scott, M. P. (1994). Five SWI/SNF gene products are components of a large multiprotein complex required for transcriptional enhancement. *Proc. Natl. Acad. Sci. USA* 91, 2905-2908.
- Phillips, M. D. and Shearn, A. (1990). Mutations in *polycomb*, a *Drosophila* Polycomb-group gene, cause a wide range of maternal and zygotic phenotypes. *Genetics* 125, 91-101.
- Pillus, L. and Grunstein, M. (1995). Chromatin structure and epigenetic regulation in yeast. In *Chromatin structure and gene expression*, Elgin, S. C. R., ed. (Oxford University Press), pp. 123-146.
- Pirrotta, V., Chan, C. S., McCabe, D. and Qian, S. (1995). Distinct parasegmental and imaginal enhancers and the establishment of the expression pattern of the *Ubx* gene. *Genetics* 141, 1429-1450.

References

- Pirrotta, V. and Rastelli, L. (1994). *white* gene expression, repressive chromatin domains and homeotic gene regulation in *Drosophila*. *Bioessays* 16, 549-556.
- Platero, J. S., Hartnett, T. and Eissenberg, J. C. (1995). Functional analysis of the chromodomain of HP1. *EMBO J.* 14, 3977-3986.
- Platero, J. S., Sharp, E. J., Adler, P. N. and Eissenberg, J. C. (1996). *In vivo* assay for protein-protein interactions using *Drosophila* chromosomes. *Chromosoma* 104, 393-404.
- Poole, S. J., Kauvar, L. M., Drees, B. and Kornberg, T. (1985). The *engrailed* locus of *Drosophila*: Structural analysis of an embryonic transcript. *Cell* 40, 37-43.
- Poux, S., Kostic, C. and Pirrotta, V. (1996). Hunchback-independent silencing of late *Ubx* enhancers by a Polycomb group response element. *EMBO J.* 15, 4713-4722.
- Pultz, M. A., Diederich, R. J., Cribbs, D. L. and Kaufman, T. C. (1988). The *proboscipedia* locus of the Antennapedia complex: a molecular and genetic analysis. *Genes Dev.* 2, 901-920.
- Qian, S., Capovilla, M. and Pirrotta, V. (1991). The *bx* region enhancer, a distant *cis*-control element of the *Drosophila Ubx* gene and its regulation by *hunchback* and other segmentation genes. *EMBO J.* 10, 1415-1425.
- Qian, S., Capovilla, M. and Pirrotta, V. (1993). Molecular mechanisms of pattern formation by the BRE enhancer of the *Ubx* gene. *EMBO J.* 12, 3865-3877.
- Raff, J. W., Kellum, R. and Alberts, B. (1994). The *Drosophila* GAGA transcription factor is associated with specific regions of heterochromatin throughout the cell cycle. *EMBO J.* 13, 5977-5983.
- Rastelli, L., Chan, C. S. and Pirrotta, V. (1993). Related chromosome binding sites for *zeste*, suppressors of *zeste* and Polycomb group proteins in *Drosophila* and their dependence on Enhancer of *zeste* function. *EMBO J.* 12, 1513-1522.
- Ray, R. P. and Schüpbach, T. (1996). Intercellular signalling and the polarisation of body axes during *Drosophila* oogenesis. *Genes Dev.* 10, 1711-1723.
- Reijnen, M. J., Hamer, K. M., den Blaauwen, J. L., Lambrechts, C., Schoneveld, I., van Driel, R. and Otte, A. P. (1995). Polycomb and *bmi-1* homologues are expressed in overlapping patterns in *Xenopus* embryos and are able to interact with each other. *Mech. Dev.* 53, 35-46.

References

- Renauld, H., Aparicio, O. M., Zierath, P. D., Billington, B. L., Chhablani, S. K. and Gottschling, D. E. (1993). Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and by *SIR3* dosage. *Genes Dev.* 7, 1133-1145.
- Reuter, G. and Spierer, P. (1992). Position effect variegation and chromatin proteins. *Bioessays* 14, 605-612.
- Riley, P. D., Carroll, S. B. and Scott, M. P. (1987). The expression and regulation of *Sex combs reduced* protein in *Drosophila* embryos. *Genes Dev.* 1, 716-730.
- Rivera-Pomar, R., Lu, X., Perrimon, N., Taubert, H. and Jäckle, H. (1995). Activation of posterior gap gene expression in the *Drosophila* blastoderm. *Nature* 376, 253-256.
- Rivier, D. H. and Pillus, L. (1994). Silencing speaks up. *Cell* 76, 963-966.
- Sabl, J. F. and Henikoff, S. (1996). Copy number and orientation determine the susceptibility of a gene to silencing by nearby heterochromatin in *Drosophila*. *Genetics* 142, 447-458.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning: A laboratory manual. Second edition (Cold Spring Harbor Laboratory Press).
- Sanchez-Herrero, E., Vernós, I., Marco, R. and Morata, G. (1985). Genetic organisation of *Drosophila* bithorax complex. *Nature* 313, 108-113.
- Santamaria, P. and Randsholt, N. B. (1995). Characterisation of a region of the X chromosome of *Drosophila* including *multi sex combs (mx)*, a Polycomb group gene which also functions as a tumour suppressor. *Mol. Gen. Genet.* 246, 282-290.
- Sathe, S. S. and Harte, P. J. (1995). The *Drosophila extra sex combs* protein contains WD motifs essential for its function as a repressor of homeotic genes. *Mech. Dev.* 52, 77-87.
- Sato, T., Hayes, P. H. and Denell, R. E. (1984). Homeosis in *Drosophila*: Maternal effect of the *Enhancer of Polycomb* locus and its interaction with *Polycomb* and related loci. *Dev. Genet.* 4, 185-198.
- Sato, T., Russell, M. A. and Denell, R. E. (1983). Homeosis in *Drosophila*: a new enhancer of *Polycomb* and related homeotic mutations. *Genetics* 105, 357-370.
- Schedl, P. and Grosveld, F. (1995). Domains and boundaries. In *Chromatin structure and gene expression*, Elgin, S. C. R., ed. (Oxford University Press), pp. 172-196.

References

- Schlossher, J., Eggert, H., Paro, R., Cremer, S. and Jack, R. S. (1994). Gene inactivation in *Drosophila* mediated by the *Polycomb* gene product or by position effect variegation does not involve major changes in the accessibility of the chromatin fibre. *Mol. Gen. Genet.* 243, 453-462.
- Schneider, I. (1972). Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morphol.* 27, 353-365.
- Schumacher, A., Faust, C. and Magnuson, T. (1996). Positional cloning of a global regulator of anterior-posterior patterning in mice. *Nature* 383, 250-253.
- Scott, M. P. and Weiner, A. J. (1984). Structural relationships among genes that control development: sequence homology between the *Antennapedia*, *Ultrabithorax* and *fushi tarazu* loci of *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 81, 4115-4119.
- Simon, J., Bornemann, D., Lunde, K. and Schwartz, C. (1995). The extra sex combs product contains WD40 repeats and its time of action implies a role distinct from other Polycomb group products. *Mech. Dev.* 53, 197-208.
- Simon, J., Chiang, A. and Bender, W. (1992). Ten different *Polycomb* group genes are required for spatial control of *abdA* and *AbdB* homeotic products. *Development* 114, 493-505.
- Simon, J., Chiang, A., Bender, W., Shimell, M. J. and O'Connor, M. (1993). Elements of the *Drosophila* bithorax complex that mediate repression by *Polycomb* group products. *Dev. Biol.* 158, 131-144.
- Simon, J., Peifer, M., Bender, W. and O'Connor, M. (1990). Regulatory elements of the bithorax complex that control expression along the anterior-posterior axis. *EMBO J.* 9, 3945-3956.
- Sinclair, D. A. R., Campbell, R. B., Nicholls, F., Slade, E. and Brock, H. W. (1992). Genetic analysis of the *Additional sex combs* locus of *Drosophila melanogaster*. *Genetics* 130, 817-825.
- Smouse, D., Goodman, C., Mahowald, A. and Perrimon, N. (1988). *polyhomeotic*, a gene required for the embryonic development of axon pathways in the central nervous system of *Drosophila*. *Genes Dev.* 2, 830-842.
- Soeller, W. C., Oh, C. E. and Kornberg, T. B. (1993). Isolation of cDNAs encoding the *Drosophila* GAGA transcription factor. *Mol. Cell. Biol.* 13, 7961-7970.

References

- Solomon, M. J., Larsen, P. L. and Varshavsky, A. (1988). Mapping protein-DNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene. *Cell* 53, 937-947.
- Solomon, M. J. and Varshavsky, A. (1985). Formaldehyde-mediated DNA-protein crosslinking: a probe for *in vivo* chromatin structures. *Proc. Natl. Acad. Sci. USA* 82, 6470-6474.
- St Johnston, D. and Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. *Cell* 68, 201-219.
- Strahl-Bolsinger, S., Hecht, A., Luo, K. and Grunstein, M. (1997). SIR2 and SIR4 interactions differ in core and extended telomeric heterochromatin in yeast. *Genes Dev.* 11, 83-93.
- Struhl, G. (1981a). A gene product required for correct initiation of segmental determination in *Drosophila*. *Nature* 293, 36-41.
- Struhl, G. (1981b). A homeotic mutation transforming leg to antenna in *Drosophila*. *Nature* 292, 635-638.
- Struhl, G. (1982). Genes controlling segmental specification in the *Drosophila* thorax. *Proc. Natl. Acad. Sci. USA* 79, 7380-7384.
- Struhl, G. (1983). Role of the *esc+* gene product in ensuring the selective expression of segment-specific homeotic genes in *Drosophila*. *J. Embryol. exp. Morph.* 76, 297-331.
- Struhl, G. (1984). Splitting the bithorax complex of *Drosophila*. *Nature* 308, 454-457.
- Struhl, G. and Akam, M. (1985). Altered distributions of *Ultrabithorax* transcripts in *extra sex combs* mutant embryos of *Drosophila*. *EMBO J.* 4, 3259-3264.
- Struhl, G. and Brower, D. (1982). Early role of the *esc+* gene product in the determination of segments in *Drosophila*. *Cell* 31, 285-292.
- Struhl, G., Struhl, K. and Macdonald, P. M. (1989). The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell* 57, 1259-1273.
- Struhl, G. and White, R. A. H. (1985). Regulation of the *Ultrabithorax* gene of *Drosophila* by other bithorax complex genes. *Cell* 43, 507-519.
- Strutt, D. I. and White, R. A. H. (1994). Characterisation of T48, a target of homeotic gene regulation in *Drosophila* embryogenesis. *Mech. Dev.* 46, 27-39.

References

- Strutt, H. L., Cavalli, G. and Paro, R. (1997). Co-localisation of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J. in press*.
- Studier, F. W. and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high level expression of cloned genes. *J. Mol. Biol.* 189, 113-130.
- Tabata, T., Schwartz, C., Gustavson, E., Ali, Z. and Kornberg, T. B. (1995). Creating a *Drosophila* wing de novo, the role of *engrailed*, and the compartment border hypothesis. *Development* 121, 3359-3369.
- Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufmann, T. C. and Kennison, J. A. (1992). *brhma*: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* 68, 561-572.
- Tautz, D. (1988). Regulation of the *Drosophila* segmentation gene *hunchback* by two maternal morphogenetic centres. *Nature* 332, 281-284.
- Thompson, J. S., Ling, X. and Grunstein, M. (1994). The histone H3 amino terminus is required for telomeric and silent mating locus repression in yeast. *Nature* 369, 245-247.
- Thüringer, F., Cohen, S. M. and Bienz, M. (1993). Dissection of an indirect autoregulatory response of a homeotic *Drosophila* gene. *EMBO J.* 12, 2419-2430.
- Tiong, S. Y. K. and Russell, M. A. (1990). Clonal analysis of segmental and compartmental homeotic transformations in *Polycomb* mutants of *Drosophila melanogaster*. *Dev. Biol.* 141, 306-318.
- Treich, I., Cairns, B. R., de los Santos, T., Brewster, E. and Carlson, M. (1995). SNF11, a new component of the yeast SNF-SWI complex that interacts with a conserved region of SNF2. *Mol. Cell. Biol.* 15, 4240-4248.
- Tripoulas, N., LaJeunesse, D., Gildea, J. and Shearn, A. (1996). The *Drosophila ash1* gene product, which is localised at specific sites on polytene chromosomes, contains a SET domain and a PHD finger. *Genetics* 143, 913-928.
- Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G. and Reuter, G. (1994). The protein encoded by the *Drosophila* position-effect variegation suppressor gene *Su(var)3-9* combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J.* 13, 3822-3831.

References

- Tsukiyama, T., Becker, P. B. and Wu, C. (1994). ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* 367, 525-532.
- Tsukiyama, T., Daniel, C., Tamkun, J. and Wu, C. (1995). *ISWI*, a member of the *SWI2/SNF2* ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* 83, 1021-1026.
- Tsukiyama, T. and Wu, C. (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* 83, 1011-1020.
- Tsukiyama, T. and Wu, C. (1997). Chromatin remodelling and transcription. *Curr. Opin. Genet. Dev.* 7, 182-191.
- Udvardy, A., Maine, E. and Schedl, P. (1985). The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J. Mol. Biol.* 185, 341-358.
- van der Lugt, N. M. T., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., van der Valk, M., Deschamps, J., Sofroniew, M., van Lohuizen, M. and Berns, A. (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *bmi-1* proto-oncogene. *Genes Dev.* 8, 757-769.
- van Holde, K., Zlatanova, J., Arents, G. and Moudrianakis, E. (1995). Elements of chromatin structure: histones, nucleosomes, and fibres. *In* Chromatin structure and gene expression, Elgin, S. C. R., ed. (Oxford University Press), pp. 1-26.
- van Lohuizen, M., Frasch, M., Wientjens, E. and Berns, A. (1991). Sequence similarity between the mammalian *bmi-1* proto-oncogene and the *Drosophila* regulatory genes *Psc* and *Su(z)2*. *Nature* 353, 353-355.
- Wall, G., Varga-Weisz, P. D., Sandaltzopoulos, R. and Becker, P. B. (1995). Chromatin remodeling by GAGA factor and heat shock factor at the hypersensitive *Drosophila hsp26* promoter *in vitro*. *EMBO J.* 14, 1727-1736.
- Wallrath, L. L. and Elgin, S. C. R. (1995). Position effect variegation in *Drosophila* is associated with an altered chromatin structure. *Genes Dev.* 9, 1263-1277.

References

- Wallrath, L. L., Lu, Q., Granok, H. and Elgin, S. C. R. (1994). Architectural variations of inducible eukaryotic promoters: preset and remodeling chromatin structures. *Bioessays* 16, 165-170.
- Walter, J., Dever, C. A. and Biggin, M. D. (1994). Two homeo domain proteins bind with similar specificity to a wide range of DNA sites in *Drosophila* embryos. *Genes Dev.* 8, 1678-1692.
- Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R. and Crabtree, G. R. (1996). Diversity and specialisation of mammalian SWI/SNF complexes. *Genes Dev.* 10, 2117-2130.
- Wedeen, C., Harding, K. and Levine, M. (1986). Spatial regulation of *Antennapedia* and bithorax gene expression by the *Polycomb* locus in *Drosophila*. *Cell* 44, 739-748.
- White, R. A. H. and Lehmann, R. (1986). A gap gene, *hunchback*, regulates the spatial expression of *Ultrabithorax*. *Cell* 47, 311-321.
- Wilson, C. J., Ghao, D. M., Imbalzano, A. N., Schnitzler, G. R., Kingston, R. E. and Young, R. A. (1996). RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell* 84, 235-244.
- Wines, D. R., Talbert, P. B., Clark, D. V. and Henikoff, S. (1996). Introduction of a DNA methyltransferase into *Drosophila* to probe chromatin structure *in vivo*. *Chromosoma* 104, 332-340.
- Winston, F. and Carlson, M. (1992). Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *TIG* 8, 387-391.
- Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A. J. and Korsmeyer, S. J. (1995). Altered *Hox* expression and segmental identity in *Mll*-mutant mice. *Nature* 378, 505-508.
- Zavortink, M. and Sakonju, S. (1989). The morphogenetic and regulatory functions of the *Drosophila Abdominal-B* gene are encoded in overlapping RNAs transcribed from separate promoters. *Genes Dev.* 3, 1969-1981.
- Zecca, M., Basler, K. and Struhl, G. (1995). Sequential organising activities of engrailed, hedgehog and decapentaplegic in the *Drosophila* wing. *Development* 121, 2265-2278.
- Zhang, C.-C. and Bienz, M. (1992). Segmental determination in *Drosophila* conferred by hunchback (*hb*), a repressor of the homeotic gene Ultrabithorax (*Ubx*). *Proc. Natl. Acad. Sci. USA* 89, 7511-7515.

References

- Zhang, C.-C., Müller, J., Hoch, M., Jäckle, H. and Bienz, M. (1991). Target sequences for *hunchback* in a control region conferring *Ultrabithorax* expression boundaries. *Development* 113, 1171-1179.
- Zhao, K., Hart, C. M. and Laemmli, U. K. (1995). Visualisation of chromosomal domains with boundary element-associated factor BEAF-32. *Cell* 81, 879-889.
- Zhou, J., Barolo, S., Szymanski, P. and Levine, M. (1996). The Fab-7 element of the bithorax complex attenuates enhancer-promoter interactions in the *Drosophila* embryo. *Genes Dev.* 10, 3195-3201.
- Zink, B., Engström, Y., Gehring, W. J. and Paro, R. (1991). Direct interaction of the *Polycomb* protein with *Antennapedia* regulatory sequences in polytene chromosomes of *Drosophila melanogaster*. *EMBO J.* 10, 153-162.
- Zink, B. and Paro, R. (1989). *In vivo* binding pattern of a *trans*-regulator of homeotic genes in *Drosophila melanogaster*. *Nature* 337, 468-471.
- Zink, D. and Paro, R. (1995). *Drosophila Polycomb*-group regulated chromatin inhibits the accessibility of a *trans*-activator to its target DNA. *EMBO J.* 14, 5660-5671.

- APPENDIX -

A1. Restriction endonuclease sites in the bithorax complex

Total sequence of BX-C (SEQ89E): 338234 bp

Enzyme	Site	(fragment length) coordinates					
EcoRI	G'AATTC	(9)	9	(744)	753	(633)	1386
		(2260)	3646	(3612)	7258	(422)	7680
		(10319)	17999	(347)	18346	(1149)	19495
		(5588)	25083	(942)	26025	(32)	26057
		(526)	26583	(14190)	40773	(7024)	47797
		(923)	48720	(3810)	52530	(568)	53098
		(736)	53834	(5375)	59209	(236)	59445
		(2672)	62117	(1668)	63785	(1130)	64915
		(111)	65026	(381)	65407	(98)	65505
		(1568)	67073	(3434)	70507	(5894)	76401
		(1927)	78328	(1353)	79681	(4344)	84025
		(89)	84114	(4389)	88503	(1637)	90140
		(2170)	92310	(3802)	96112	(1637)	97749
		(2935)	100684	(5242)	105926	(660)	106586
		(3102)	109688	(5989)	115677	(150)	115827
		(3052)	118879	(2588)	121467	(2305)	123772
		(7652)	131424	(5316)	136740	(674)	137414
		(6741)	144155	(1447)	145602	(49)	145651
		(2088)	147739	(2899)	150638	(1890)	152528
		(1894)	154422	(165)	154587	(641)	155228
		(295)	155523	(3060)	158583	(1361)	159944
		(3870)	163814	(4941)	168755	(745)	169500
		(164)	169664	(15828)	185492	(596)	186088
		(378)	186466	(2257)	188723	(2489)	191212
		(974)	192186	(675)	192861	(2210)	195071
		(3115)	198186	(2716)	200902	(6909)	207811
		(8113)	215924	(2317)	218241	(3384)	221625
		(1023)	222648	(1320)	223968	(1131)	225099
		(1646)	226745	(891)	227636	(4883)	232519
		(6807)	239326	(3901)	243227	(3314)	246541
		(842)	247383	(6200)	253583	(1768)	255351
		(4108)	259459	(3985)	263444	(388)	263832
		(7480)	271312	(375)	271687	(1614)	273301
		(2571)	275872	(6154)	282026	(2333)	284359
		(511)	284870	(5675)	290545	(279)	290824
		(584)	291408	(9768)	301176	(10074)	311250

Appendix

		(3355)	314605	(1852)	316457	(7026)	323483
		(831)	324314	(7048)	331362	(4288)	335650
		(1809)	337459	(775)			
HindIII	A'AGCTT	(3013)	3013	(6623)	9636	(2982)	12618
		(1646)	14264	(6396)	20660	(6473)	27133
		(4212)	31345	(11908)	43253	(1464)	44717
		(11841)	56558	(448)	57006	(976)	57982
		(846)	58828	(246)	59074	(5507)	64581
		(4071)	68652	(111)	68763	(6313)	75076
		(4436)	79512	(3041)	82553	(3612)	86165
		(1571)	87736	(7816)	95552	(7421)	102973
		(13467)	116440	(3317)	119757	(15305)	135062
		(3266)	138328	(1437)	139765	(8928)	148693
		(14537)	163230	(1188)	164418	(2133)	166551
		(1398)	167949	(13456)	181405	(10768)	192173
		(4216)	196389	(2444)	198833	(11879)	210712
		(2257)	212969	(1144)	214113	(758)	214871
		(6722)	221593	(532)	222125	(20085)	242210
		(2469)	244679	(5142)	249821	(1972)	251793
		(7367)	259160	(1487)	260647	(2332)	262979
		(10018)	272997	(2392)	275389	(11863)	287252
		(9304)	296556	(28)	296584	(1707)	298291
		(10569)	308860	(1101)	309961	(16796)	326757
		(925)	327682	(5671)	333353	(4198)	337551
		(683)					
SalI	G'TCGAC	(19488)	19488	(3658)	23146	(11644)	34790
		(984)	35774	(8380)	44154	(8836)	52990
		(7155)	60145	(97)	60242	(40644)	100886
		(6)	100892	(632)	101524	(2487)	104011
		(1019)	105030	(3033)	108063	(2001)	110064
		(4449)	114513	(10424)	124937	(3835)	128772
		(4497)	133269	(2851)	136120	(2308)	138428
		(3708)	142136	(11166)	153302	(305)	153607
		(5733)	159340	(2487)	161827	(11098)	172925
		(5899)	178824	(7511)	186335	(237)	186572
		(7)	186579	(7)	186586	(10795)	197381
		(24675)	222056	(8106)	230162	(9360)	239522
		(8645)	248167	(11208)	259375	(17719)	277094
		(467)	277561	(571)	278132	(3173)	281305
		(7472)	288777	(2823)	291600	(465)	292065
		(3617)	295682	(17859)	313541	(2076)	315617
		(5779)	321396	(16838)			

Appendix

XhoI	CTCGAG	(26250)	26250	(23760)	50010	(3326)	53336
		(14895)	68231	(1866)	70097	(11066)	81163
		(11501)	92664	(11905)	104569	(11523)	116092
		(3908)	120000	(2259)	122259	(5566)	127825
		(2201)	130026	(5002)	135028	(4858)	139886
		(724)	140610	(606)	141216	(7745)	148961
		(936)	149897	(1081)	150978	(16018)	166996
		(7728)	174724	(22138)	196862	(415)	197277
		(1089)	198366	(25483)	223849	(128)	223977
		(933)	224910	(1127)	226037	(5955)	231992
		(5412)	237404	(3554)	240958	(13)	240971
		(30896)	271867	(1513)	273380	(37)	273417
		(4434)	277851	(3076)	280927	(55)	280982
		(3289)	284271	(6269)	290540	(1800)	292340
		(5004)	297344	(7552)	304896	(2835)	307731
		(181)	307912	(4342)	312254	(6023)	318277
		(10581)	328858	(5325)	334183	(4051)	

A2. Restriction endonuclease sites in putative PREs of the bithorax complex

Figure A1 shows a detailed restriction map and the sites of known transcripts and regulatory elements in the PC binding fragments

Peak A (9636 - 12618 HindIII fragment)

Total sequence: 2982 bp

No restriction sites for BamHI, EcoRI, PstI, SalI, XhoI

Enzyme	Site	(fragment length) coordinates					
AccI	GTMKAC	(278)	278	(1136)	1414	(912)	2326
		(656)					
AseI	ATTAAT	(810)	810	(1127)	1937	(329)	2266
		(487)	2753	(202)	2955	(27)	
RsaI	GTAC	(84)	84	(30)	114	(422)	536
		(465)	1001	(1051)	2052	(379)	2431
		(551)					

Peak B (109688 - 115677 EcoRI fragment)

Total sequence: 5989 bp

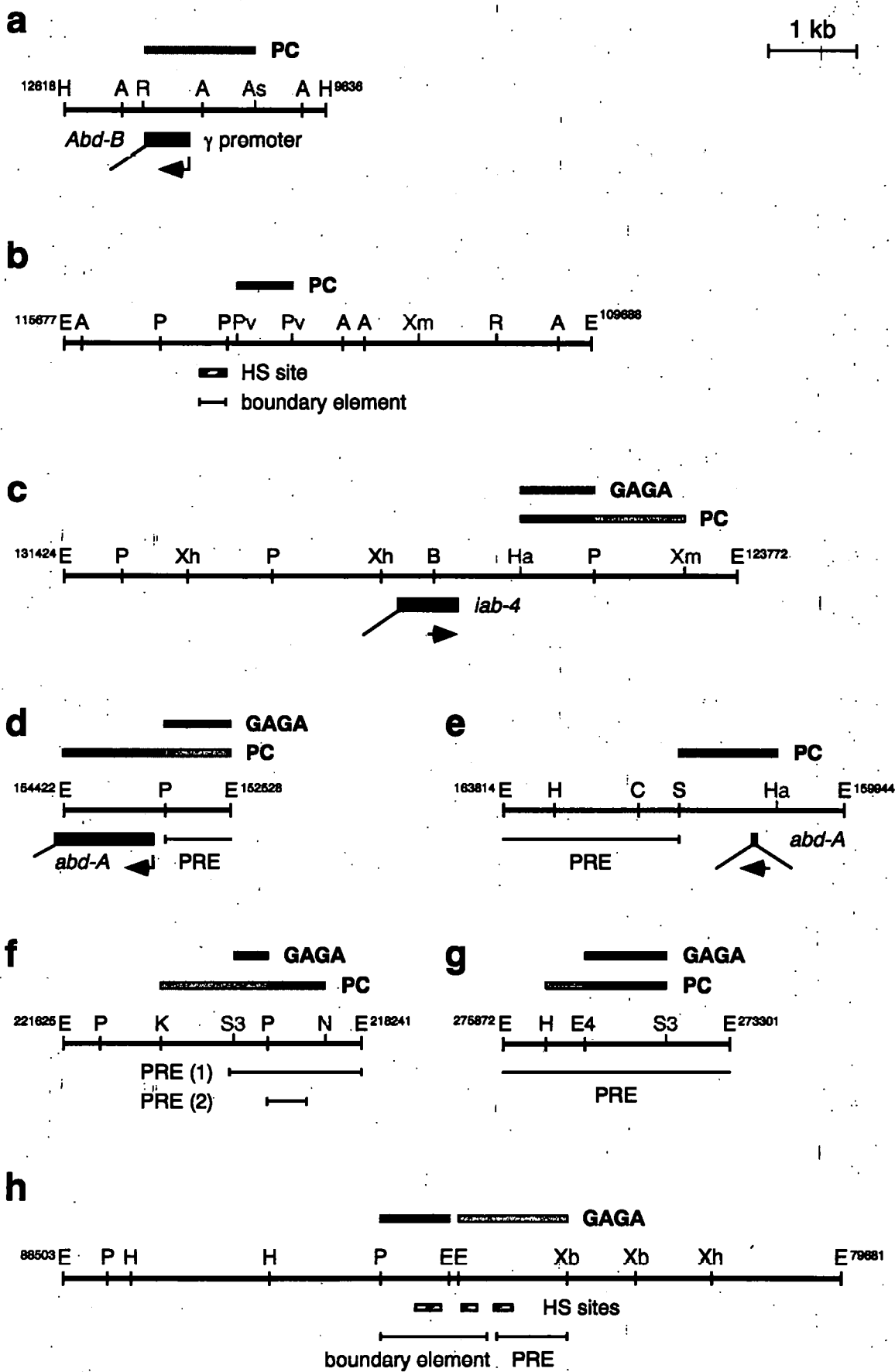
No restriction sites for BamHI, HindIII, XhoI

Figure A1. Correlation of the Polycomb and GAGA factor binding elements with the genetic elements of the bithorax complex

Panels (a) - (g) describe the PC binding elements A - G respectively, and panel (h) shows the *Fab-7* domain. Restriction maps of the subclones used in figures 3.3, 3.4 and 3.7 are shown (see also Appendix A2), with proximal to the left and distal to the right. The coordinates of the sites used for subcloning are given, based on the map of Martin et al., 1995. Restriction enzymes marked are AccI (A), AseI (As), BamHI (B), ClaI (C), EcoRI (E), Eco47III (E4), HaeII (Ha), HindIII (H), KpnI (K), NdeI (N), PstI (P), PvuII (Pv), RsaI (R), SalI (S), Sau3A (S3), XbaI (Xb), XhoI (Xh), XmnI (Xm).

The dark grey bars above each map show the peak binding elements for PC and/or GAGA factor, as described in figures 3.4 and 3.7. Regions of high binding adjacent to the peak domains are shown in lighter grey. Shown below each map are:-

- (a) The position of the *Abd-B* γ promoter and first exon, together with the direction of transcription (Martin et al., 1995).
- (b) The position of a prominent DNaseI hypersensitive site, and the minimal deletion defining the *Mcp* boundary element (Karch et al., 1994).
- (c) The second exon of the non-coding *iab-4* transcript (Cumberledge et al., 1990).
- (d) The position of the *abd-A* promoter and first exon (Martin et al., 1995), together with the *iab-3* PRE (Simon et al., 1993; Chiang et al., 1995).
- (e) The position of a 50 bp *abd-A* exon (Martin et al., 1995), and the *iab-2* PRE (Simon et al., 1993; Chiang et al., 1995).
- (f) The position of the *bxd* PRE, as described by Chan et al., 1994 (1), and a smaller element (2) mediating Polycomb responses in transient tissue culture experiments (Chang et al., 1995).
- (g) The *bx* PRE (Simon et al., 1993; Qian et al., 1993; Chiang et al., 1995).
- (h) The positions of three DNaseI hypersensitive sites within the *Fab-7* regulatory element (Karch et al., 1994), and of the *Fab-7* boundary element and the *iab-7* PRE (Hagstrom et al., 1996).



Appendix

Enzyme	Site	(fragment length) coordinates					
AccI	GT ⁺ MKAC	(377)	377	(2206)	2583	(2243)	2826
		(889)	5715	(274)			
PstI	CTGCA'G	(4140)	4140	(754)	4894	(1095)	
PvuII	CAG'CTG	(3405)	3405	(623)	4028	(867)	4895
		(1094)					
RsaI	GT'AC	(1071)	1071	(2817)	3888	(747)	4635
		(316)	4951	(31)	4982	(813)	5795
		(194)					
SalI	G'TCGAC	(376)	376	(4449)	4825	(1164)	
XmnI	GAANN'	(1963)	1963	(4008)	5971	(18)	
	NNTTC						

Peak C (123772 - 131424 EcoRI fragment)

Total sequence: 7652 bp

No restriction sites for HindIII

Enzyme	Site	(fragment length) coordinates					
BamHI	G'GATCC	(3447)	3447	(4205)			
HaeII	RGCGC'Y	(1256)	1256	(1211)	2467	(5185)	
PstI	CTGCA'G	(1625)	1625	(3665)	5290	(1704)	6994
		(658)					
SalI	G'TCGAC	(1165)	1165	(3835)	5000	(2652)	
XhoI	C'TCGAG	(4053)	4053	(2201)	6254	(1398)	
XmnI	GAANN'	(691)	691	(1114)	1805	(4035)	5840
	NNTTC	(1686)	7526	(126)			

Peak D (152528 - 154422 EcoRI fragment)

Total sequence: 1894 bp

No restriction sites for BamHI, HindIII, XhoI

Enzyme	Site	(fragment length) coordinates					
PstI	CTGCA'G	(773)	773	(1121)			
SalI	G'TCGAC	(774)	774	(305)	1079	(815)	

Peak E (159944 - 163814 EcoRI fragment)

Total sequence: 3870 bp

No restriction sites for BamHI, PstI, XhoI

Enzyme	Site	(fragment length) coordinates				
ClaI*	ATCGAT	(2336)	2336	(1534)		
HaeII	RGCGC'Y	(765)	765	(1257)	2022	(1848)
HindIII	A'AGCTT	(3286)	3286	(584)		
SaII	G'TCGAC	(1883)	1883	(1927)		

* Published sequence (Martin et al., 1995) reports an additional ClaI site at position 1662.

Peak F (218241 - 221625) EcoRI fragment

Total sequence: 3384 bp

No restriction sites for BamHI, Sall, XhoI

Enzyme	Site	(fragment length) coordinates				
HindIII	A'AGCTT	(3352)	3352	(32)		
KpnI	GGTAC'C	(2286)	2286	(1098)		
NdeI	CA'TATG	(413)	413	(2971)		
PstI	CTGCA'G	(1077)	1077	(1903)	2980	(404)
Sau3A	'GATC	(1008)	1008	(450)	1458	(1414)
		(115)	2987	(397)	2872	

Peak G (273301 - 275872 EcoRI fragment)

Total sequence: 2571 bp

No restriction sites for BamHI, Sall

Enzyme	Site	(fragment length) coordinates				
Eco47III	AGC'GCT	(1659)	1659	(912)		
HindIII	A'AGCTT	(2088)	2088	(483)		
PstI	CTGCA'G	(2416)	2416	(155)		
Sau3A	'GATC	(725)	725	(1708)	2433	(27)
		(111)			2460	
XhoI	C'TCGAG	(79)	79	(37)	116	(2455)

Fab-7 clone 1 (79681 - 84025 EcoRI fragment)

Total sequence: 4344 bp

No restriction sites for Sall, PstI

Enzyme	Site	(fragment length) coordinates				
BamHI	G'GATCC	(738)	738	(3606)		
HindIII	A'AGCTT	(2872)	2872	(1472)		
XbaI	T'CTAGA	(2336)	2336	(778)	3114	(1230)
XhoI	C'TCGAG	(1482)	1482	(2862)		

Fab-7 clone 2 (84114 - 88503 EcoRI fragment)

Total sequence: 4389 bp

No restriction sites for BamHI, Sall, XhoI

Enzyme	Site	(fragment length) coordinates				
HindIII	A'AGCTT	(2051)	2051	(1571)	3622	(767)
PstI	CTGCA'G	(782)	782	(3105)	3887	(502)

A3. Restriction endonuclease sites in the *emgrailed-injected* genomic walk

A detailed restriction map of the *en-inv* walk is shown in figure A2.

